The Demonstrative Model of First-Person Thought

Daniel Morgan, UCL

Suppose I think "I am hungry". The object of that thought will be myself. Why is that? What determines the reference of our first-person thoughts? On the view defended here, our ways of gaining of knowledge of ourselves do. We have a multitude of very rich ways of gaining knowledge of ourselves – e.g. proprioception, kinaesthesia. The view defended is that these ways of gaining knowledge are what enable us to latch onto ourselves in first-person thought. An analogous view about the mechanisms for reference-determination is highly plausible for some other kinds of thought. Suppose I am looking directly at a bottle placed before me on a table. I think "That bottle is small". We may stipulate that I do not possess any name that refers to the bottle, and do not use any descriptions it uniquely satisfies to single it out. Nevertheless, I can in fact single the bottle out. Intuitively, that is because vision – a very rich way of gaining knowledge of objects in one's environment – enables me to latch onto it in thought. On the view about first-person thought defended here, it is similarly the case that when I think a first-person thought, my various ways of gaining knowledge of the object my thought is about – myself – are what enable me to latch onto that object in thought.

Just so as to have a label, I shall refer to the view I defend as *The Demonstrative Model* of first-person thought. The label is natural since the view implies that the story about how the reference of first-person thoughts is determined is parallel to the most natural story to tell about how the reference of at least one kind of thought that can be expressed using a demonstrative expression – our example is a visual demonstrative thought – is determined. But it is no part of what *The Demonstrative Model* claims that "I" is really a demonstrative expression (perhaps meaning the same as "this body" or "this mind"), or that first-person thoughts are really a kind of demonstrative thought. No assimilation of one kind of thought to a different kind of thought is intended. Relatedly, it would not matter for my purposes if it somehow turned out that, actually, the right story to tell about how the reference of thoughts like "That bottle is small" is determined does not mention ways of gaining knowledge. My focus here is just on the question of how the reference of first-person thoughts is determined. The comparison between first-person thoughts, as I want to conceive of them, and demonstrative thoughts, as it highly natural to conceive of these, is purely an expository device.

In section 1, I introduce the principle that I believe supplies the best motivation for *The Demonstrative Model*. This is the knowledge-maximization principle, which says that assignments of reference ought to maximize knowledge. The knowledge-maximization principle is controversial, and it needs to be clarified in various ways. But a good initial case for it can be made. With the principle in place, we have a way of adjudicating questions about how the reference of various kinds of thoughts is determined. I argue that, in the case of first-person thoughts, the principle adjudicates in favor of *The Demonstrative Model*.

Gareth Evans argued for a version of *The Demonstrative Model* in *The Varieties of Reference*.[1] However, most subsequent theorists of first-person thought have regarded *The Demonstrative Model* as a non-starter. So any defender of the model will have to put considerable effort into what are primarily defensive maneuvers. The remaining sections of the paper defend *The Demonstrative Model* against what I take to be the most influential objections to it.

In section 2, I consider a strategy to show that *The Demonstrative Model* is at best redundant. The basic thought behind the strategy is that there is an alternative, less contentious model available of what determines the reference of first-person thoughts.[2]This alternative model – *The Simple Rule Model* – says that first-person thoughts are governed by the simple rule that any token of one will be about the person whose token it is, i.e. the person who produces the token. If *The Demonstrative Model* is inconsistent with *The Simple Rule Model*, the argument goes, then it should be rejected. But if it is consistent with *The Simple Rule Model*, then it is redundant. So *The Demonstrative Model* is at best redundant.

I will argue that there is a robust and a non robust way of understanding what *The Simple Rule Model* says, corresponding to robust and non robust way of understanding the notion of a rule the model invokes. On the robust way of understanding *The Simple Rule Model* what the model says is false. On the non-robust way of understanding it, what the model says is true. But it fails to address the explanatory question addressed by *The Demonstrative Model.*

In section 3, I rebut the argument that there are referential guarantees on first-person thought – e.g. the guarantee that any attempt to think a first-person thought will succeed in netting an object – that *The Demonstrative Model* fails to respect.[3] I do this not by denying that any of the putative guarantees exists – an option tried out by Evans – but by explaining how *The Demonstrative Model* can respect the guarantees.

In section 4, I consider some more miscellaneous objections to *The Demonstrative Model*, and I offer replies.

Knowledge-Maximization and The Demonstrative Model (I)

> Knowledge-Maximization Principle: assignments of reference should maximize knowledge.

Timothy Williamson defends the knowledge-maximization principle.[4] His basic form of argument is inference to the best explanation. He describes cases in

---

[1] I lack space to look at Evans's version in detail here. One way in which his version differs from mine is in its motivation. Evans motivates his model not by reference to the knowledge-maximization principle but by reference to his requirement that thinking a singular thought always involves knowing which thing one is thinking about. There are obscurities both in relation to how this requirement is to be understood, and in what the argument for it is. For effective criticism, see Manley and Hawthorne (2012).

[2] This strategy is central in Campbell (1994). The idea that first-person thoughts are governed by a simple rule is also prominent in O Brien (2007) and Peacocke (2008)

[3] This kind of argument is central in Campbell (1994), and developed further in O Brien (2007). One of the thought experiments the argument relies on go back to Anscombe (1975).

[4] See Williamson (2007)

which it is clear what the intuitively correct assignment of reference to an utterance or thought is. Then he argues that the knowledge-maximization principle does a better job of predicting the intuitively correct assignment than rival principles.

In one of Williamson's cases, a subject, Emmanuel, has a mistaken belief that he can discern facts about a person's character and life-history just by looking at their face. He sees a stranger, Celia, standing some distance away (2007: 263). He judges "She is F, G, H….", ascribing a character and life-history in considerable detail. None of it fits Celia. By pure coincidence some of it does fit Elsie, whom Emmanuel has never seen or heard of.

Why is it that Emmanuel's beliefs are about Celia and not about Elsie? 'Because Emmanuel can see the Celia but not Elsie' is a perfectly good answer to this question. But there is a natural follow-up question: why is it that Emmanuel's beliefs are about the woman he can see, and not about the woman whom he cannot see but who fits more of his beliefs? Why does reference go with vision in this way?

If one endorses a Davidsonian truth-maximization principle – a principle that says that assignments of reference should maximize true belief – rather than a knowledge-maximization principle, this may be difficult to explain. An assignment of reference on which Emmanuel's beliefs are beliefs about Elsie maximizes the number of true beliefs he is credited with. However, that assignment will not maximize knowledge (since all of the beliefs Emmanuel formed about character and life-history were based on an unreliable method, none of them constitute knowledge). If one accepts the knowledge-maximization principle, one therefore has an explanation of why reference goes with vision in this case. Emmauels's beliefs *are not* about Elsie because, if his beliefs were about her, although many of his beliefs would be true, none of them would be knowledge. Emmanuel's beliefs *are* about Celia because, if they are about her, at least some of them will be knowledge. For example, Emmanuel's perceptual judgment "She is in front of me" will be knowledge.

Of course, the truth-maximization principle is not the only rival to the knowledge-maximization principle. There are also principles according to which causal links are what determines reference. Vision is a causal link. So a rival account of what is going on is that Emmanuel refers to Celia and not Elsie because there is a causal link between him and Celia, but not one between him and Elsie.

Williamson's response is that not all causal links are channels for reference. For example, there will be causal links between Emmanuel and countless causal intermediaries on the causal chain between his visual experience and Celia. None of these causal links is a channel for reference. The causal links that are channels for reference, Williamson says, are precisely those that are also channels for knowledge.

Suppose for the moment that this kind of argument for the knowledge-maximization principle is persuasive. Suppose that the reason that seeing an object can enable one to latch onto that object in thought is that seeing is a way

of gaining knowledge. Are there any lessons to draw for an account of how the reference of first-person thoughts is determined? Consider the following case. I am proprioceiving the position of my own body and I think "I have crossed knees". On an assignment of myself as the reference of my first-person thought, my thought will count as knowledge, since proprioception can put me in a position to know that my knees are crossed. Thus, the *explanation* of how vision enables reference in the visual demonstrative case – vision is a way of gaining knowledge, and assignments of reference maximize knowledge – seems to apply *mutatis mutantis* in the case of the proprioception-based first-person thought we are looking at. The explanation is: proprioception is a way of gaining knowledge, and assignments of reference maximize knowledge.

Of course, this does not constitute a water-tight case for *The Demonstrative Model*. For one thing, *The Demonstrative Model* makes a claim about how the reference of first-person thoughts in general is determined. The case looked at was a particularly favorable kind of case for *The Demonstrative Model*. We will see in later sections what *The Demonstrative Model* can say about first-person thoughts that are not knowledge.

In addition, the case made is conditional on something like the knowledge-maximization principle being correct.[5] Seriously to defend that principle would require more space than I have here.[6] But it will at least be helpful to say a little more about how I think maximization principles like the knowledge-maximization principle ought to be understood.

---

[5] Why "something like"? Well, suppose one thought that reference should be assigned so as to maximize justified belief rather than knowledge, or that reference should be assigned so as to maximize justified true belief rather than knowledge. A parallel explanation of how proprioception enables reference could still be given. Proprioception is a way of gaining justified belief, and of gaining justified true belief, just as much it is a way of gaining knowledge. Moreover, in cases in which one fails to have knowledge of someone who nevertheless happens to "fits" one's beliefs, one will often also fail to have justified beliefs, or justified true beliefs, about them. E.g., Emmanuel lacks justification for beliefs about the character and life-history of Elsie just as much as he lacks knowledge of her. What this brings out is that the defender of *The Demonstrative Model* need not buy into the distinctive "knowledge first" approach Williamson pursues. So the "something like" really does considerably reduce the contentiousness of the motivation.

To keep the discussion manageable though, I will focus for the moment on motivating *The Demonstrative Model* just by reference to the knowledge-maximization principle. Having seen, in later sections of the paper, more of the defense of *The Demonstrative Model* (in particular, the discussion of quasi-proprioception in section III) it should be clearer what range of principles can serve as motivations for *The Demonstrative Model*. For now, I simply signal that I think that a justified belief maximization principle, or a justified true belief maximization principle, would do. But a truth-maximization principle would not do. Neither would a principle that just appeals to causal links, such as McGlynn (2011) defends. Neither would a principle according to which there is nothing substantive and general one can say about how thoughts get their references, such as Martin(2009) expresses sympathy for.

[6]McGlynn (2011) and Martin (2009) both aim to undermine knowledge-maximization by describing a wider variety of problem cases than Williamson considers. McGlynn attempts to describe a case in which causal links play a role in fixing reference even when they do not enable knowledge. Martin attempts to describe a case in which it is possible to selectively attend to one rather than the other of two objects despite the fact that one is in an equally good position to gain knowledge of either of them. Fans of something like the knowledge-maximization principle other than Williamson include Brewer (2000) and Campbell (1994). Campbell's sympathy is particularly relevant since he is probably the most influential critic of *The Demonstrative Model.*

Clarifying Maximization Principles

To maximize something is to create as much as possible of that thing. Possible given what? Given whatever one is taking as fixed – whatever one has as one's constraints. What constraints are supposed to apply to the knowledge-maximization principle? Williamson's own answer to this is not totally clear. His proposal, he says, "is to replace true belief by knowledge in a principle of charity constitutive of content" (2007:264). The principle of charity she discusses is the Davidsonian one according to which a radical interpreter should interpret those he seeks to understand as having as much true belief as possible. So one might think that the knowledge-maximization principle Williamson endorses is the principle that a radical interpreter should interpret those she seeks to understand as havng as much knowledge as possible. But Williamson also expresses approval of Lewis's remark (1974:110-11) that the predicament of the radical interpreter is "merely a literary device for dramatizing the question: how do the intentional states of agents supervene on the non-intentional states of the world?" (2007:260). Thus it is natural to infer that the knowledge-maximization principle Williamson endorses is the principle that intentional facts supervene on non-intentional facts in such a way as to maximize knowledge. Assignments of reference should create as much knowledge as possible, *taking the non-intentional facts as fixed*.

One problem with this interpretation is that Williamson elsewhere in the same chapter expresses an open mind on whether the principles relating the intentional and the non intentional do in fact fix the supervenience of the former on the latter (2007:261).

A different problem with the interpretation is that the explanation of an individual case of reference-determination of Williamson's that we looked at seemed to take some intentional facts for granted. Williamson's explanation of why Emmanuel's thought is about Celia is that (i) Emmanuel can see Celia and (ii) seeing is a way of gaining knowledge.

But the fact that seeing is a way of gaining knowledge is an intentional fact – it is a fact about knowledge. It could be pointed out that is also plausibly a necessary fact, if it is a fact at all. But why should that matter?

Perhaps more seriously, the fact that Emmuel can see Celia is only a non-intentional fact on the controversial assumption that visual experiences lack intentional content. Williamson seems not to want to commit himself to this assumption (2007:263). But how can it be right to think of the kind of explanation he offers as an explanation of how, in a particular case, the intentional supervenes on the non-intentional if the explanation takes for granted that Emmanuel can see Celia, and if the Emmanuel's seeing Celia consists in the fact fact that Emmanuel has a visual experience with the intentional property of being about Celia?

There is an even simpler problem. On the assumption that visual experiences are never themselves states of knowledge – as opposed to being the grounds for

such states – there is no way of applying the knowledge-maximization principle to them. So there is no way of using knowledge-maximization to explain how visual experiences get their intentional contents.

What these points show, I think, is that billing the knowledge-maximization principle as a principle about how intentional facts supervene on non-intentional facts makes the task of defending that principle unduly demanding. In particular, with the principle construed in this way, Williamson's examples do not seem to provide even a prima facie justification for the knowledge-maximization principle.

Nevertheless Williamson's examples are illuminating of something, and they can be used to defend a less ambitious knowledge-maximization principle. Minimally, the point of a knoweldge-maximization principle should be to allow us to say something substantive about why various kinds of thoughts have the objects that they do. Given that motivation, we can legitimately appeal to any facts that aren't facts about what objects thoughts have. E.g., we can appeal to which objects a subject can see (or proprioceive), even if facts about which objects a subject can see (or proprioceive) are intentional facts. Interpreting the knowledge-maximization principle in this way may undermine its use in gaining mileage against what Williamson calls "judgment scepticism", which is a part of what Williamson himself wants the principle to do.[7] But it does not undermine its use in adjudicating between competing accounts of how the reference of thoughts of a particular kind – e.g. first-person thoughts – is determined.

(II) The Demonstrative Model and the Simple Rule Model

We can use an intuitive contrast between first-person thoughts and visual demonstrative thoughts as a way of introducing and motivating the dominant alternative to *The Demonstrative Model*.

The intuitive contrast is this. Even in advance of any particular story about how vision enables us to latch onto objects in thought – e.g. in advance of a story that appeals to knowledge-maximization – we know that this is something vision does. We can describe cases in which there is really nothing but the fact that a subject can see an object that could explain how they are latching onto that object in thought – cases in which the subject has no name to pick out the object, and does not rely on any description in picking it out. Merely taking as fixed that a particular thought is a thought of a kind that can be expressed by a demonstrative expression, and that it is the thought of a particular subject, nothing at all follows about which object the thought will be about. So there is always an explanatory question about why the thought has the object that it does, and not some different object. In some cases, vision – or perhaps some more demanding vision-involving relationship, like the relationship of visual attention – will be all that one could reasonably appeal to in answering that question.

Contrast this with the case of first-person thought. Merely taking for granted that a particular thought can be expressed by the first-person, and that it is a thought

[7] See 2007: 275. See also McGlynn(2011) for discussion of whether Williamson is right to think that knowledge-maximization would help with judgment skepticism.

of a particular subject, something does follow about which object the thought will be about. It follows that the thought will be about the subject themselves. This follows from what many have taken to be the obvious fact that any token first-person thought will be about the person whose thought it is, i.e. the person who produces the token. Since this fact alone determines who a particular first-person thought will be about, there is no *need* for, for example, proprioception to do the kind of singling out work that vision does in the case of visual demonstrative thoughts.

The metaphorical picture we have, in the case of a visual demonstrative, is that of a first arrow shooting out at an object – this arrow represents vision. The first arrow is then followed by a second arrow – this arrow represents thought. The second arrow tracks the first. But, in the first-person case, we already know what trajectory the arrow representing thought must describe. It must describe a loop. So, the suggestion is, there is no need to regard the arrow representing thought as tracking the progress of any first arrow – e.g. one representing proprioception. If there happens also to be a first arrow that describes a loop that is interesting. But we could say what the right trajectory for the second arrow is without reference to it, or to anything else.

John Campbell is the most influential defender of the idea that the fact that any first-person thought will be about the subject undermines the need for any appeal to ways of gaining knowledge in an explanation of how the reference of first-person thoughts is determined. He writes:

> Only with reluctance should we take this step, which can seem so easy and indeed inescapable, of supposing that there is a use of the first person on which its reference is fixed not by the token-reflexive rule [the rule that any token of the first person will be about the producer of that token] but in some other way. Many of the most distinctive phenomena involving the first person are straightforwardly explained by its being governed by the rule. Once we leave the rule behind, problems swarm upon us. Most immediately and simple is the question of whether the alternative method of reference-fixing, whatever it is, is guaranteed to yield the same results as the token-reflexive rule in what references it finds for particular uses of the first person. If there is this guaranteed coincidence, then it is not apparent what advantage there can be in shifting to the new method. If the coincidence is not guaranteed, then we have opened up the possibility of someone using the first person to refer to someone other than himself. But this would not be recognizable as a use of the ordinary first person. (Campbell, "Past, Space and Self" p. 124)[8]

[8]Equally trenchant quotations to similar effect appear in most of Campbell's more recent papers on the first-person. For example:

> "If the first-person has a reference, it is specified by a simple rule, the token-reflexive rule: 'Any token of "I" refers to whoever produced it'. This is the most straightforward way of stating how the reference of "I" is fixed. In fact, I would argue that it is really the only way of saying how the reference of the first-person is fixed. It never goes wrong: there are no significant counterexamples to this as an account of the reference of "I".

The token-reflexive rule (or "simple rule", as Campbell also calls it, and as I shall follow him in calling it) Campbell explicitly speaks of here applies to the first person – a particular linguistic device. But it is clear that Campbell thinks that something analogous holds for first-person thoughts. The analogous simple rule for first-person thoughts is the rule that any token first-person thought will be about the producer of that token. Campbell in fact allows the categories of thought and language to blur somewhat. But it is important, if we are properly to evaluate the choice between *The Demonstrative Model* and *The Simple Rule Model* – the model that says that the reference of first-person thoughts is fixed by the rule that any token first-person thought will be about the producer of that token – that we not allows these two categories to blur.[9]

The claim that the word "I" is governed by the simple rule is fairly plausible (though it is by no means uncontroversial). It is fairly plausible that the word "I" is governed by the simple rule because it is fairly plausible that the word "I" is governed by a *convention* whose content is the content of the simple rule. Rules in language just are conventions.

The parallel suggestion, at the level of thought, would be that first-person thoughts are governed by the convention that any token first-person thought token will be about the subject who produces that token. But that just seems wrong. Thoughts, unlike words, are not typed purely syntactically. Thoughts do not get their objects in virtue of being governed by conventions. To be a first-person thought is already to have a certain semantic profile.

It might be replied that this objection assumes an overly robust notion of a rule. On a less robust understanding, to claim that there is a rule to the effect that any token first-person thought will be about the subject who produces that token is not to invoke a convention or anything similar. It is just to claim that it is invariably correct to say that any token first-person thought will be about the subject whose thought it is.

At this point though, the defender of *The Demonstrative Model* can turn the tables on *The Simple Rule Model*. Campbell's thought was that, if an alternative reference-fixing story to *The Simple Rule Model* is compatible with *The Simple Rule Model*, then there "no advantage to shifting to the new method". But, on the non robust way of understanding *The Simple Rule Model* we have been forced to consider, this seems wrong. The point of *The Demonstrative Model* is to explain why first-person thoughts have the references that they do. If we agree that it is invariably correct to say that any token first-person thought will be about the subject who produces that token, then the task for *The Demonstrative Model* will be to explain why it is that any token first-person thought will be about the

---

> Alternative accounts of how the reference of "I" is fixed are satisfactory only insofar as they agree with the determination of the token-reflexive rule." 2012:6

[9]What makes it a little surprising that Campbell allows the two categories to blur is that it seems an important part of the position of Evans, the best known defender of *The Demonstrative Model*, that language and thought wok a bit differently – see my fn 13. The difference between language and thought that I will to appeal to here however is not one that Evans mentions.

subject who produces that token. So, *The Demonstrative Model* will aim to explain precisely what *The Simple Rule Model* takes for granted. If *The Demonstrative Model* is successful in this aim it should be preferred to *The Simple Rule Model.* It should be preferred even if both models make exactly the same predictions about which things "particular uses of the first person" refer to.

The Simple Rule Model and Knowledge-Maximization

We should consider a final way of defending *The Simple Rule Model*. Suppose we are offered the following picture of how the reference of first-person thoughts is determined. We take as primitive the fact that all the different token first-person thoughts there are are tokens of a single semantic type. The question we ask ourselves is what semantic profile (or "rule") should be assigned to this semantic type.

In answering this question we appeal to knowledge-maximization. We say that the right semantic profile to assign to the first-person thought type is the rule that any token of the first-person thought type is about the person who produces that token. The reason that this is the right rule is that it is the rule that maximizes the number of first-person thoughts that are knowledge. And, of course, a major part of the explanation for why that is the rule that maximizes the number of first-person thoughts that are knowledge will consist in facts like the fact that proprioception is a way of gaining knowledge of oneself, but not, or at least not usually, of things other than oneself.

Someone with this picture – call it *The Simple Rule Model Plus* – is already conceding quite a lot to *The Demonstrative Model*. The claim *The Demonstrative Model* makes is that our ways of gaining knowledge of the objects of our first-person thoughts – ourselves – play a role in determining their reference, just as our ways of gaining knowledge of the objects of our demonstrative thoughts – the objects in our environment – play a role in fixing their reference. On *The Simple Rule Model Plus*, this claim seems true, not false. So *The Simple Rule Model Plus* is a version of *The Demonstrative Model*.

Even so, it is not a version of *The Demonstrative Model* we should accept. The argument against it I wish to offer requires us to think about the different levels at which the knowledge-maximization principle might be applied.[10]

[10]A different way of arguing against *The Simple Rule Model Plus* – one which I will not attempt to explore properly here – is to put pressure on the notion of a *producer* the simple rule invokes. Clearly, that notion is a causal notion. But, equally clearly, there are cases in which someone who is in some sense the cause of a particular first-person thought token is not the object of that thought token. If such cases are not to be taken as counterexamples to the simple rule a rather refined notion of producer is called for. One possibility – a possibility that Campbell (1994) actually expresses an open mind about – is that the relevant notion of producer is a partly epistemic notion. In hard cases – e.g. in the hard cases that dominate the literature on personal identity – we work out who the producer of a given thought token is by working out which person the subject is in a position to gain knowledge about in characteristic first-person ways. On this picture, we already need to consult facts about which object one is in a position to gain knowledge of just to work out what verdict the simple rule issues in relation to a particular case. In that case though, it is hard to see why one would bother appealing to a simple rule as opposed to just appealing directly to the reference-determining potential of ways of gaining knowledge.

One might apply the knowledge-maximization principle to individual token thoughts in a completely atomistic way, without assuming anything about how those token thoughts relate to other token thoughts. Or one might apply the principle to a set of token thoughts, all of which are to be assigned the same reference. Or one might apply the principle, as *The Simple Rule Model Plus* does, to a set of token thoughts, taking as fixed that all those token thoughts are of the same semantic type, and are all governed by the same semantic rule, even if they are not all necessarily to be assigned the same reference.

The first option looks to be a non-starter. Some thoughts are not knowledge. For those that aren't, applying the knowledge-maximization principle in a completely atomistic way will never explain why they have the reference that they have. On the correct assignment for such thoughts, the amount of knowledge generated will be no knowledge. So the correct assignment will never generate any more knowledge than any other assignment.

The second option looks to be Williamson's practice in the Emmanuel case. Williamson asks what person should be assigned as the reference of Emmanuel's various visual demonstrative thoughts – his well-founded thought about Celia's being in front of him, as well as his entirely fanciful thoughts about Celia's character and life-history. Williamson does not consider any assignments on which Emmanuel's different visual demonstrative thoughts have different objects. He takes for granted that they are all to be assigned the same object.

What justifies this? One thing we can say to justify it is that Emmanuel himself will treat his different thoughts about Celia as having a single object. For example, he will be prepared to perform inferences like "She is in front of me. She is paranoid. So, someone paranoid is in front of me." It seems a point in favor of an assignment if it assigns a single object to thoughts that the thinker of the thoughts treats as having a single object. That this is a point in favor of an assignment might even have its ultimate justification in knowledge-maximization. After all, inferences comprised of thoughts that the subject of the inference treats as all having a single object will only actually extend knowledge if those thoughts are assigned a single object.

If we have this preference for assignments that assign just a single object to thoughts that the thinker of the thoughts treats as having a single object, we can equally appeal to it to explain how the reference of first-person thoughts that are not knowledge is determined. Suppose I judge and know, on the basis of proprioception, "My knees are crossed". Suppose I judge but do not know "I am a great poet". On the basis of these two thoughts together, I conclude "I am a great poet with knees crossed". Here, the right person to assign as the reference of the second premise and of the conclusion is myself. An atomistic application of the knowledge-maximization principle has no explanation for this. But we can explain it as follows. The knowledge-maximizing assignment for my judgment "My knees are crossed" is the assignment on which that judgment is about me, the object I can proprioceive. This, together with the fact that I treat the second premise and the conclusion of the inference as being about the same thing as the

first premise is about, is what explains why the second premise and the conclusion are also both about me. We might call this the "inferential glue" explanation of how first-person thoughts that are not knowledge get their reference determined.

There is an alternative to the inferential glue explanation, which is offered by *The Simple Rule Model Plus*. According to that model, our task is to assign a rule to the first-person thought type. Cases of first-person thoughts where things are going well epistemically – e.g. my "My knees are crossed" thought, based on proprioception – are appealed to in support of the claim that the right rule to assign to the first-person type, granted that a single rule is assumed to govern all tokens of that type, is the token-reflexive rule. The token-reflexive rule then itself explains why, for example, my "I am a great poet with knees crossed" thought is about me. The thought is about me because it is of the first-person type, and I am its producer. That is different from saying, as the inferential glue explanation does, that the thought is about me because I treat it in inference as being about the same object as another thought that has been determined independently to be about me.

I think a version of *The Demonstrative Model* that incorporates the inferential glue explanation is preferable to *The Simple Rule Model Plus* for two reasons.

First, a version of *The Demonstrative Model* that incorporates the inferential glue explanation does not require us to accept as an unexplained primitive that all the tokens of the first-person type are tokens of the same semantic type.[11] So it is explanatorily more powerful than *The Simple Rule Model Plus.*

Second, as we have already seen, the difficulty raised by the existence of thoughts that are not knowledge for a defender of the knowledge-maximization principle has nothing to do with first-person thoughts in particular. It equally arises for, for example, Emmanuel's visual demonstrative thoughts. But the solution to that problem, as it arises in the first-person case, that is defended by *The Simple Rule Model Plus* does not readily generalize beyond the first-person case. Suppose we assume that all token thoughts that one might express with "she" are tokens of the same semantic type, and try to work out what rule to assign to that type. We might say that the rule is that any "she"-thought will be about the female person demonstrated. But even if this were correct – and it at least needs to be refined to deal with "she"-thoughts whose reference is determined anaphorically rather than demonstratively – it does not seem as though we have enough independent handle on how to individuate demonstrations to work out when the reference of two "she"-thoughts is fixed by

[11] Someone might try to argue that *The Simple Rule Model Plus* need not simply take this for granted. We are justified in taking all the relevant tokens as being of a single semantic type, it might be said, by the fact that they are all tokens of the same physical type. But offering this justification is incompatible with what seems to be a perfectly standard kind of multiple realization possibility – the possibility that the physical types that realize different peoples' tokenings of the first-person type are different.

a single demonstration – when there is a constraint that they be assigned the same reference – and when there is no such constraint.[12]

From this perspective, it looks like a quirk that in the case of the first person there may be a way of refraining from imposing a preference for assignments on which thoughts that are treated by their subjects as having a single object really do have a single object. We need to impose that preference anyhow to deal with cases not involving the first person. Given that we do, there is no extra cost in appealing to it as well in the first person case. And, as we have seen, there is an explanatory gain.

On examination, the initially stark-seeming opposition between *The Demonstrative Model* and *The Simple Rule Model* becomes fuzzy at the edges. The best version of *The Simple Rule Model* – *The Simple Rule Model Plus* – is a version of *The Demonstrative Model*, one that is distinguished by giving a simple rule a modest explanatory role, and by taking as primitive the fact that all token first-person thoughts are of a single semantic type. Even so, this is not the best version of *The Demonstrative Model*. The best version of *The Demonstrative Model* is the one that relies on the inferential glue explanation, makes no appeal to any simple rule, and does not need to take as primitive the fact that all token first-person thoughts are of a single semantic type.

(III) Referential Guarantees on First-Person Thought and The Demonstrative Model

The last section took for granted that *The Demonstrative Model* could make the right predictions about which things particular first-person thoughts are about – that the model was at least extensionally correct. It focused on the question of whether, assuming the model's extensional correctness, there was reason to prefer it to *The Simple Rule Model*. But it is in fact highly controversial whether *The Demonstrative Model* is even extensionally correct. In this section, I counter arguments that it is not.

Lucy O Brien locates *The Demonstrative Model* in the work of Gareth Evans:

> It has, following the work of neo-Fregeans such as Evans and Peacocke, come to seem, at least to other Fregeans, relatively unproblematic to suppose that our account of demonstrative reference will draw heavily on the perceptual relations that obtain between the referrer and the object referred to. In the case of demonstrative reference such perceptual information can be seen as providing us with an epistemological basis that will serve in an explanation of how demonstrative reference is accomplished….It is not clear that there is the same possibility of being in receipt of information of a kind that will serve to identify the reference of a use of the first person. The challenge that seems to face the Fregean is to give an account of the epistemological basis that can underwrite first person reference in the way that perceptual information underwrites

[12] I recognize that a convincing defense of the hunch expressed in this paragraph would need to consider a wider variety of strategies for formulating the rule governing "she"-thoughts. E.g. an alternative strategy might focus on the notion of salience, rather than that of a demonstration.

> much demonstrative reference….This is what Evans aims to do in *The Varieties of Reference*. He offers an account of the information channels that are in operation, that will, he claims, underwrite first person reference. "I" is treated along the lines of other demonstratives (O Brien, "Evans on Self-Identification" p. 233).

O Brien pithily sums up what she takes to be the major problem for *The Demonstrative Model* as follows:

> "I" has a special propensity to refer and to stick with its object of reference, even though the waves of information might seem to want to buffet it this way and that. ("Evans on Self-Identification" p. 241)

O Brien's observation gestures at two kinds of putative problem case for *The Demonstrative Model*, which she goes on to discuss. The first kind of case is one in which one is wired up so that experiences of a sort that are generally caused by the condition of one's own body are in fact caused by the condition of someone else's body – e.g. a case of quasi-proprioceptive experience in which I am wired up to someone else's knees. The second kind of case is the Anscombe-inspired case in which many of one's ordinary routes to knowledge of oneself are suspended, because one is inside a sensory deprivation chamber. I will deal with the two kinds of problem case in turn.

The Quasi-Proprioception Case

Why might the quasi-proprioception case seem problematic for *The Demonstrative* Model? Well, in a case of quasi-proprioception, one is receiving information – at least in some minimal causal sense of the term 'information' – not from one's own body but from someone else's. But the key thought behind *The Demonstrative Model* is that information plays a reference-determining role. So, it might seem, the waves of information should buffet one's first-person thought towards the person one is actually gaining information from. But, as O Brien says, this seems like the wrong result. There is a guarantee that any first-person thought will be about oneself, not about anybody else.

The defender of *The Demonstrative Model* can say that their model allows a reference-determining role only to one specific kind of causal-informational channel – the kind that is a channel for knowledge. The question then arises of whether quasi-proprioception is a channel of this kind. The defender of *The Demonstrative Model* may point out that if one judges "My knees are crossed" on the basis of a quasi-proprioceptive experience whose cause is the condition of someone else's body, one's judgment will not be knowledge since the person one's judgment is about – the subject – is different from the person one is actually in a position to gain knowledge of.

This reply misses the point of the objection. It takes for granted that any first-person judgment will be about the subject. But this is something *The Demonstrative Model* is supposed to explain, not take for granted. The objection was precisely that, in this particular case, *The Demonstrative Model* cannot explain why one's first-person thought shouldn't be about the person from whom one is receiving proprioceptive experience. After all, the thought is, the

subject *is* in a position to gain knowledge of someone, just not themselves. So if ways of gaining knowledge really do fix the reference of first-person thoughts this is a case where the reference of the first-person thought should be to someone other than the subject.

The defender of *The Demonstrative Model* needs to deny that when one has quasi-proprioceptive experiences caused by the state of someone else's body a channel for knowledge exists between oneself and that person. This is a controversial but still highly defensible position.[13] What is true – or can be stipulated – is that a reliable connection between the subject and the other person's knees exists. But that doesn't prove that the subject is in a position to gain knowledge of the other person's knees. Given the latitude there is in how processes are individuated, any belief at all can be seen as the result of a reliable belief-forming process. So the notion of reliability does not provide us with an independent way of adjudicating the question of whether a channel for knowledge exists.

We need to address that question directly by asking whether if a subject in the quasi-proprioception case judges, just on the basis of the experience they have, "Somebody's knees are crossed", their judgment will be knowledge. There are two cases to distinguish. If the subject makes that judgment because it follows from the judgment "My knees are crossed", then the existentially general judgment inherits its epistemic credentials from the first-person judgment. Since the first-person judgment is not knowledge, neither is the existentially general judgment. On the other hand, the subject might make the judgment for no reason at all – as a complete "shot in the dark", to use Evans's phrase. It that case though, their judgment is irrational, and fails to be knowledge for that reason.

Admittedly, there is a third possibility to consider. The subject might judge "Somebody's knees are crossed" because they suspect that their proprioceptive experience is – or even just might be – caused by the condition of someone else's body. Suppose they know that this kind of thing is common in their environment. In this case, their judgment may well be knowledge. But it will be knowledge they have arrived at from an inference containing premises that mention their experience and its possible aetiology. It is not knowledge they have arrived just by relying on their experience, and its being knowledge does nothing to show that quasi-proprioception itself can be a channel for knowledge of the state of someone else's knees. Similarly, if I learn that a particular kind of visual hallucination – say that of there being a dagger being before me – is reliably caused by the presence of a certain object – say, a harmless kitchen knife – in my

[13]Similarly, it is defensible to deny that a channel for justified belief, or a channel for justified true belief, exists. This is one reason why I said, in fn 5, that the defender of *The Demonstrative Model* only needs "something like" the knowledge-maximization principle to be correct. A justified belief maximization principle, or justified true belief maximization principle, would also do. But notice that a Davidsonian truth-maximization principle – i.e. one on which true belief is what is maximized – would not do. That is because the only obstacle to the subject's having a true belief when he judges "My knees are crossed" is the fact that, intuitively, the subject's belief is about himself, not about the person who is the cause of his experiences. So, the defender of a truth-maximization principle has no explanation of why the subject's belief is not about the person who is the cause of his experiences.

environment I may be able to use my knowledge that I am having an experience as of a dagger to work out whether there is a kitchen knife around. But that does not make the hallucination of the dagger a channel for reference to the kitchen knife.[14]

The Sensory Deprivation Chamber

Why might the sensory deprivation chamber case seem problematic for a defender of *The Demonstrative Model*? Well, in a sensory deprivation chamber, one is receiving very little if any information about one's own body. So it might seem that one has no basis for knowledge of oneself (we can stipulate if we like that one is at the same time suffering from amnesia, and therefore has no memory based knowledge). So it might seem that *The Demonstrative Model* will predict that in this kind of case singling oneself out in first person thought will be impossible. This seems like the wrong result. It seems as though one will at least be able to *attempt* to think a first-person thought in the sensory deprivation chamber. And – unlike in the case of visual demonstrative thoughts – it seems there is a guarantee that any attempt to think a first-person thought will succeed in netting an object. Evans denied the existence of this guarantee. He argued that it was based on an unjustified inference from the observation that any attempt to use the word "I" to refer was bound to succeed in referring.[15] But following Evans in this denial feels like a cost.

The defender of *The Demonstrative Model* should instead insist that even in the sensory deprivation, even suffering from amnesia, one has sources for knowledge of oneself. For example, one can introspect.[16] So plenty of one's first-person thoughts about one's mental properties will be knowledge. Admittedly, there will be other first-person thoughts one may think that, in virtue of being in a sensory deprivation chamber, will not be knowledge. For example, if one

---

[14] For discussion of the issues quasi-proprioceptive experiences raise, see the essays in Recanati and Prosser (eds) (2012) on the notion of immunity to error through misidentification. It is interesting to note that this is a point at which Evans's defense of *The Demonstrative Model* seems to be helped by his deflationary position about the significance of the possibility of quasi-proprioceptive experiences. Evans thought that judgments like "My knees were crossed", based on proprioception, were immune to error through misidentification despite that possibility.

[15] Discussing the sensory deprivation case, Evans writes: "This idea [the idea that 'whenever there is a subject, then there is at least one thing he can unproblematically think about, namely himself'] comes, I think, from *a false analogy with the functioning of "I" as a communicative device*. For it seems perfectly true that provided there is a speaker, he can *refer to* himself, using the device of self-reference "I"….But what this means is that it is always possible for an audience to think of the referent of a token of such a device as the person who uttered that token (1982: 252)". My italics.

[16] Appealing to introspection marks a difference between the version of *The Demonstrative Model* I defend and the version Evans defends. Evans does not seem to allow introspection, as opposed to our ways of gaining knowledge about our bodily properties, any role in determining the reference of our first-person thoughts. From the point of view of the knowledge-maximization principle though, ignoring introspection would be arbitrary, given that it is a *bona fide* way of gaining knowledge. If one didn't want to appeal to introspection here though – e.g. because one didn't believe any such faculty existed – one might instead appeal to one's way of gaining knowledge of the future by forming intentions about one's future behavior. E.g., if I think while in the sensory deprivation chamber, "I'm not going to let this get me down", that judgment could be knowledge.

without the slightest justification thinks "My knees are crossed", then that thought will not be knowledge. But we have already seen how the defender of *The Demonstrative Model* should say how first-person thoughts that are not knowledge get their reference – by invoking the inferential glue explanation. That explanation applies inside the sensory deprivation chamber just as well as it does outside of it.

(IV) Two More Objections and Replies

Objection 1: The claim *The Demonstrative Model* makes is that our ways of gaining knowledge of the objects of our first-person thoughts help fix the reference of those thoughts. But, if the knowledge-maximization principle is true, it will be true for just any kind of thought that our ways of gaining knowledge of the objects of those thoughts help fix their reference. So what *The Demonstrative Model* says about first-person thoughts doesn't capture anything even slightly distinctive about first-person thoughts. And it doesn't present a picture of first-person thoughts on which first-person thoughts are any more like demonstrative thoughts than they, for example, like the thoughts expressed by names. But surely it is an essential component of the model that Campbell, O Brien and others were attacking that first-person thoughts are in some way especially like demonstrative thoughts.

Reply: The aim of *The Demonstrative Model* is not to capture a respect in which first-person thoughts are distinctive. It is to provide a story about what determines their reference. If the story is one that also works for other kinds of thought, so much the better. Campbell and O Brien just do say that the position they target is the position on which that our ways of gaining knowledge, or information, about ourselves are what fix the reference of our first-person thoughts. Sometimes they *also* suggest that the gist of the targeted position is that "I" means something like "this body" or "this person". But that does not seem to be an essential feature of their target. For one thing, Evans is meant to be the main defender of *The Demonstrative Model*. But Evans never said that "I" means something like "this body" or "this person".[17] And there is an easy argument that "I" does not mean the same as "this body" or "this person". It is perfectly possible to rationally doubt "I am this body" or "I am this person". So there is at least a difference in sense here, even if there is no difference in reference. More generally, there seems to be no good reason to expect to be able to assimilate first-person thinking to any other kind of thinking.

Objection 2: The ways of gaining knowledge of oneself that have been mentioned so far – e.g. proprioception, introspection – are *reflexive* ways of gaining knowledge – i.e. ways of gaining knowledge that also provide knowledge about oneself. But often a first-person thought is knowledge in virtue of the exercise of a way of gaining knowledge that can equally provide knowledge about things other than oneself. For example, my thought "I was born in Cork" is knowledge in virtue of the operations of testimony. My parents and my passport say that I was

[17]The closest he comes to it is saying that first-person thoughts are "quasi-demonstrative". I am not totally sure what he means by saying this, but I see no reason to follow him in saying it.

born in Cork. The knowledge-maximization principle does not give reflexive ways of gaining knowledge of oneself any privileged place. So, if *The Demonstrative Model* is motivated by reference to knowledge-maximization, testimony ought be able to play the same kind of reference fixing role in relation to first-person thoughts as does, for example, proprioception. But that seems to introduce a disanalogy between first-person thought and visual demonstrative thoughts. Testimony may be what puts me in a position to know that a woman who I am looking at has just stolen a watch – someone may point to a woman we are jointly attending to and say "She stole a watch". But it seems as though vision is still responsible for enabling me to single the woman out in thought in a way that testimony is not.

Reply: Maybe testimony can play a role in fixing the reference of my "She stole a watch" thought. Maybe there are mixed visual-testimonial demonstratives. On the other hand, if there are not – if testimony can never play any reference fixing role in relation to demonstrative thoughts – then there must be some explanation of why it never can. Once that explanation has been provided, we can ask whether it extends to the first-person case.

Bibliography

Anscombe, G.E.M (1975) 'The First-Person' in Samuel Guttenpland ed., *Mind and Language* (Oxford: Clarendon Press, 1975), pp. 45-65.

Brewer, Bill (2000) 'Externalism and *A Priori* Knowledge of Empirical Facts', in C.Peacocke and P. Boghossian (eds.) *New Essays on the A Priori*: Oxford: Oxford University Press.

Campbell, J (1994) *Past, Space and Self*. Oxford: Clarendon Press.

Campbell, J (2004) 'What is it to Know What "I" Refers to?', in *The Monist* 87 (2): 206-218.

Campbell, J (2012) 'On the thesis that "I" is not a referring term' in S. Prosser and F. Recanati (eds.) *Immunity to Error Through Misidentification: New Essays*. Cambridge: Cambridge University Press.

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

Lewis, D (1974) 'Radical Interpretation'. *Synthese* 27: 331-344.

Martin, M.G.F (2009) 'Reupholstering a Discipline: Comments on Williamson', *Philosophical Studies* 145: 445-53.

Manley, D and Hawthorne, J (2012) *The Reference Book.* Oxford: Oxford University Press.

McGlynn, A (2012). 'Interpretation and Knowledge Maximization'. *Philosophical Studies* 160 (3): 391-405

O Brien, L (1995) 'Evans on self-identification'. *Nous* 29 (2): 232-247 (1995)

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

Peacocke, C (2008) *Truly Understood.* Oxford: Oxford University Press.

Prosser, S and Recanati, F (eds.) (2012) *Immunity to Error Through Misidentification: New Essays.* Cambridge: Cambridge University Press.

Williamson, T (2007). *The Philosophy of Philosophy,* Oxford: Blackwell Publishing.