The Limits of Selflessness

Semantic Relativism and the Epistemology of *De Se* Thoughts

[Penultimate draft. Forthcoming in a special issue of *Synthese*
edited by Jonas Åkerman, Peter Pagin, and Robert van Rooij]

Marie Guillot (Southampton / Institut Jean Nicod)
marie.guillot@ens.fr

Keywords: de se *thoughts; semantic relativism; (absence of) self-representation; self-location; immunity to error through misidentification; subjective perspective; centred worlds.*

Abstract
It has recently been proposed that the framework of semantic relativism be put to use to describe mental content, as deployed in some of the fundamental operations of the mind. This programme has inspired in particular a novel strategy of accounting for the essential egocentricity of first-personal or *de se* thoughts in relativist terms, with the advantage of dispensing with a notion of self-representation. This paper is a critical discussion of this strategy. While it is based on a plausible appeal to cognitive economy, the relativist theory does not fully account for the epistemic profile that distinguishes *de se* thinking, as some of its proponents hope to do. A deeper worry concerns the reliance of the theory on a primitive notion of "centre" that hasn't yet received enough critical attention, and is ambiguous between a thin and a rich reading. I argue that while the rich reading is required if the relativist analysis of the *de se* is to achieve its most ambitious aims, it also deprives the theory of much of its explanatory power.

1 *De se* thoughts: at the crossroads between epistemology and semantics

1.1 *De se* thoughts and their epistemological import

I feel cold, want to be warm, see that the heater is on my left, believe that I turned it on, wish I wouldn't mind the winter so much, imagine myself moving to sunnier climes, remember having been to Sevilla. All those thoughts[1] are first-personal or *de se* thoughts[2].

1 I will use throughout the word "thought" in the broad, Cartesian sense, to cover any type of mental state, perceptual, affective or conative as well as doxastic.

2 As the expression "first-personal thought" has the disadvantage of suggesting the involvement of a mental device of self-reference equivalent to the first-person pronoun 'I', I will prefer here the more neutral expression "*de se* thought". The notion of a "*de se*" modality of thought has been popularized by Lewis (1979) but has been in use for much longer, with first occurrences in mediaeval philosophy. Here, I use the term in a non-committing fashion, to pick out a mode of thinking that is inseparable from the thinker's subjective perspective, and *not* as shorthand for the Lewisian way of explaining the phenomenon.

De se thoughts are canonically (although not exclusively) expressed, in a subject with normal mastery of the linguistic first person and the underlying conceptual resources, by utterances containing a first-personal marker (such as the pronoun 'I'), and of the schematic self-attributive form "I am F".

The real defining feature of *de se* thoughts, however, is that they are attributions of predicates to oneself *as oneself*: they are *essentially egocentric*, i.e. they are self-involving in a non-accidental way. This is easiest to describe in terms of the first person. It is only contingent that I bear a mental relation to myself if I think something I would express by saying "The winner of the lottery is a lucky person", in a circumstance where I happen to be the winner. By contrast, as Castañeda (1968) first remarked, when I think the coextensive thought I would express by saying "I am a lucky person", my involvement cannot be accidental.

Note that the difference is epistemic: the possibility that is open in the first example, but ruled out in the second, is my *failing to be aware* that the thought concerns *me*. The property of essential egocentricity, hence the class of *de se* thoughts that it defines, are thus (at least in part) epistemic[3] phenomena. This can be put in terms made familiar by Perry (1979): whenever I have a thought which inclines me to accept sentences including first-personal markers, there is no possible paraphrase that would eliminate those indexical components without causing some loss of epistemic significance, and making the sentences unfit for psychological explanation. *De se* thoughts are, in short, "essentially indexical" (Perry 1979), just like the thoughts that dispose us to accept sentences containing "here", "now" or "this".

Let us say, then, that essential egocentricity is the egological subspecies of essential indexicality; and that *de se* thoughts are those thoughts which display essential egocentricity. *De se* thoughts, when they are true, are the repositories of self-knowledge in the strict sense, namely knowledge of ourselves *as ourselves*. Being capable of such thoughts is also what distinguishes *subjects*.

Epistemologists have further reasons for finding an interest in *de se* thoughts. The defining feature of such thoughts, namely their essential egocentricity, is, as we have seen, in large part an epistemic property. But that is only the most basic of an array of epistemic privileges that have been observed in connection to first-personal thoughts, giving them a characteristic epistemic profile.

Some of those privileges, like essential egocentricity itself, apply to all *de se* thoughts; some only apply to a core subset of those thoughts, comprised of self-ascriptions of occurrent experiences ("I'm cold", "I see the heater on my left") or of quasi-experiential, phenomenally-rich states like episodic remembering and imagining ("I'm daydreaming of Spain", "I

3 I will use the word "epistemic" in the broader way, to include what pertains, not just to knowledge but to other states relating to knowledge as to their norm – like believing, conceiving, understanding, feeling certain, being aware that *p*, etc. "Epistemic" will also be used in connection with the degree of informational richness of a subject's representations (how extensively or finely they depict the world), as this contributes to determining how much knowledge the subject possesses if those representations are true, and how much it will take to justify them. "Epistemic" is thus treated here as a rough synonym of "cognitive". The phenomena of self-location and IEM to be discussed below will be understood as epistemic in this broad sense.

remember going to Sevilla"). Self-attributions of occurrent experiences are especially reliable, and come accompanied with a typically high degree of subjective certainty. Some take those self-ascriptions to bear on states that are "transparently" or at any rate preferentially accessible to us; many accordingly claim them to be, if not incorrigible, then at least especially authoritative to a degree[4]; some consider those self-attributions to be impervious, in any case, to identity mistakes as far as their subject is concerned. (This last property, known as 'immunity to error through misidentification' or IEM for short, will be the object of Sections 2.2.3, 2.2.4 and 3.) These are just some of the various ways in which (certain) *de se* thoughts seem to be protected against error, at least to some degree: they appear to yield *qualitatively* better beliefs than the corresponding third-personal thoughts.

In addition, all *de se* thoughts go with a *quantitative* epistemic advantage, providing us with more knowledge than we would be able to gain in their absence. This is because they reflect the fact that we view reality from a certain subjective *perspective*. Thinking *de se* means representing the world in a way that also incorporates information about our own insertion in a specific part of this world. This aspect is what Lewis, Perry and Stalnaker, among others, call 'self-location'. (I will return to it in Section 2.2.1 and 2.2.2 below.)

The exact number, nature and interrelations of the features that make up the peculiar epistemic profile of first-personal thoughts have been abundantly debated. This is not the place[5] to rehearse and discuss them all; some of those features may in fact come from various sources beyond egocentric awareness *per se*, such as the nature of phenomenal consciousness. The limited purpose of this paper is to examine a family of theories of *de se* thinking that have recently been inspired by semantic relativism. To this aim, it will only be relevant to consider the last two of the epistemic privileges alluded to, namely immunity to error through misidentification and self-location. These two features have been taken by some of the relativist theories that are my object here to stem from the same source as essential egocentricity, and to be explainable, just as this latter aspect, by the nature of the semantic content these theories attribute to *de se* thoughts. But before turning to the specific explanatory project that will be at issue here, a few words must be said as to why semantics in general could be thought to be of relevance to understanding *de se* thoughts.

1.2 The semantic approach to *de se* thoughts: from self-reference to no-reference

In large part because of their typical connection with first-personal utterances, many have been tempted to link the peculiar epistemic profile of *de se* thoughts to self-reference. Underlying this approach is the following supposition: what makes all *de se* thoughts special, and secures in particular their essential egocentricity, is that they bear on a particular object – myself – envisaged from a particular angle – *as* myself. One source of the emphasis thus put on the "special access" a subject is presumed to have to herself can be traced to Frege's

4 I may of course be wrong in believing that the heater is on my left. But I can't so easily be wrong in believing that I have an experience *as of* the heater being on my left.

5 A very good survey of this debate can be found in Quassim Cassam's introduction to his (1994). Other useful sources include Alston (1971), Wright, Smith and Macdonald (1998), Gertler (2003), and Stalnaker (2008).

famous claim that "every one is presented to himself in a special and primitive way, in which he is presented to no one else. [...] And only [...] himself can grasp thoughts determined in this way."[6] Thus, a plausible source for the epistemic peculiarities of *de se* thoughts would lie in this "special and primitive", 'internal' way each person can think about *themselves*, to the exclusion of anybody else.

In addition, Fregeans take the special egocentric mode of presentation they posit to affect, not just the genesis, but also the (truth-conditional) content of *de se* thoughts. As a consequence, a given *de se* judgement (e.g. "I am cold", judged by John) will differ in its content from any coextensive third-personal judgement (e.g. "John is cold", judged by Jane). While this was certainly Frege's position, it has been more crisply expressed by Castañeda:

> No first-person proposition is identical with any third-person proposition, if both are in *oratio recta*.[7]

I will call the view expressed by this last sentence the *representational conception of the de se*, as it takes what distinguishes *de se* thinking as such to be a matter of semantic content. This has lead some to the idea that accounting for the epistemology of the *de se* devolves at least in part upon theorists of content.

In radical contrast to the above-mentioned self-reference approach, Wittgenstein (1953) and Anscombe (1975) put forward the heterodox view that *de se* thoughts should in fact be characterised by the *absence* of any (explicit) self-representation. They reject the core presupposition that the epistemic peculiarities of *de se* thinking derive from an underlying *representational achievement*, namely self-reference, marked by the use of characteristic syntactic and semantic resources. No-reference theorists remain in agreement, however, with the more general premise that a *de se* thought (like "I am cold") differs from any corresponding third-personal thought (e.g. "John is cold") on *representational grounds*. The latter refers to a particular object (John); the former, to no object at all. Both the self-reference and the no-reference views thus share the assumption that semantics can illuminate the epistemology of the *de se*.[8]

Today, the development of relativist semantics is reviving the no-reference view of the *de se* in a new guise.[9] Following forerunners such as Loar (1976), Lewis (1979), Chisholm

6 Frege (1918).

7 Castañeda (1968/1999), p. 92. Geach (1957) is a forerunner of this representationalist take on the *de se*.

8 Thus Anscombe (1975) gives a deflationary explanation of immunity to error through misidentification, based on semantic considerations: it is not, she claims, the function of a *de se* thought to represent the subject; so it cannot, *a fortiori*, *mis*represent that subject. The relativist account of IEM is in the same spirit.

9 An important *caveat* concerning this filiation. Wittgenstein and Anscombe hold a non-referential view of both first-personal thoughts and first-personal attitude reports; they explicitly claim that the pronoun 'I' does not refer. (See Wittgenstein 1953, § 404, p. 122 and § 410, p. 123; Anscombe 1975, p. 148.) By contrast, the contemporary authors discussed here put forward a non-referential treatment of the content of (a certain subclass of) *de se* attitudes, but typically remain neutral as to the relationship between this mental content and the content communicated by utterances containing a first-person pronoun. Stojanovic (2012) and Torre (2010) are rare examples of relativist analyses of *de se assertions*. But for the most part, on the linguistic side, relativist semantics has been confined to the treatment of expressions containing more diffuse marks of a subject's point of view (e.g. predicates of personal taste, PRO constructions, epistemic modals).

(1981), Sosa (1983) and Perry (1986), a number of contemporary philosophers of language and mind – among whom Egan (2006a, 2006b); Recanati (2007a, 2007b, 2010, 2012); Stephenson (2007a, 2007b, 2010); Ninan (2008, 2010)[10] – use a relativist semantics to redescribe *de se* thoughts as 'selfless' thoughts, i.e. thoughts that are not really *about* the subject who entertains them, but that are egocentric in a more essential way. My claim is that there are principled limitations to what this "selflessness" approach can achieve.

Now there are two distinct goals that a semantic theory of *de se* thoughts may aim at. The first and more modest goal is a purely descriptive one; it consists in providing a semantics for modelling the contents of *de se* thoughts. The second, more ambitious goal is an explanatory one; it consists in showing that the type of contents *de se* thoughts are claimed to have is (at least in part) responsible for (some of) their epistemological peculiarities. There is reason to take some of the contemporary relativist theories to pursue not just the first, but also the second goal.[11] My aim in this paper is to show that they are bound to fail in this respect[12]; but also that they fail in an interesting way, that may help clarify the notion of *de se* thinking.

Of course, a relativist may still insist that her objective is really the more modest, descriptive one. It would nonetheless remain an interesting question whether a theory based on semantic relativism *could* aim for the more ambitious goal of explaining *de se* epistemology, if only because of the intimate historical connection, mentioned above, between semantic and epistemological issues touching first-personal thinking. Enquiring into the prospects of a relativist theory of the more ambitious kind isn't only of polemical interest, but is useful in fixing principled bounds to the future developments of this school of thought, and in clarifying what we can or can*not* reasonably expect from it.

I start by presenting the relativist theory of *de se* thoughts, and describe how it has been applied to explain two aspects of the characteristic epistemology of those thoughts: self-location and immunity to error through misidentification (Section 2). Sections 3 and 4 are a critical discussion of the merits of this model. In Section 3, I point up two kinds of limitations in the relativist account of first-person epistemology. In Section 4, I identify a more

10 Millikan (2004) also contains related ideas.

11 Textual evidence includes Lewis's way of introducing his theory of *de se* attitudes: what a creature who entertains such attitudes has, that a creature who doesn't possess them cannot have, is a certain piece of *knowledge* (concerning their own situation within the world); and that is accounted for by the semantics. (*Cf.* the example of the two gods, presented in Section 2.2.1 below.) Egan and Ninan typically follow Lewis in this respect. Recanati, on his part, makes clear that his relativist semantics for the *de se* is dictated by considerations of psychological plausibility, and contributes to *explaining* epistemic phenomena such as immunity to error through misidentification (IEM). Here is a characteristic statement: "According to the [strong moderate relativist] account, [IEM] *follows from the fact that* (only) implicit *de se* thoughts are identification-free, since they do not involve the concept of self [...]". (2007*a*, p. 177; my emphasis.) Similar statements are to be found in Recanati (2010), pp. 484-5, Recanati (2012), p. 378, and elsewhere.

12 Let me make clear that the criticisms I will raise against the relativist theory by no means imply an endorsement of competing self-reference views. I do *not*, in particular, think that a token-reflexive approach fares better in explaining *de se* epistemology. I expect in fact that no semantic theory can do so on its own. However, the project of offering a positive account of *de se* epistemology, whether in non-purely representational terms or not, would go far beyond the scope of this paper. My (limited) aim here is only to show the specific ways in which relativism about the *de se* fails in its declared explanatory ambition (*cf.* the previous note), assessing that ambition on its own terms. These shortcomings are sufficiently distinctive (and different from the limitations of self-reference views) to deserve attention on their own.

fundamental source of concern: the theory's reliance on a notion of perspective or "centre" which, when examined more closely, turns out to be richer than is usually acknowledged. This suggests a general diagnosis for the various shortcomings of the theory. I draw two consequences from this diagnosis. First, given the substantiveness of the primitives that the relativist, "selfless" theory of egocentricity must presuppose to achieve its more ambitious aims, it is not clear to what extent the view is really more economical that the competing self-representational approaches. Second, and for the same reason, the theory's contribution to our understanding of the *de se*, while significant, may well turn out to be more descriptive than explanatory. However persuasively a relativist theory of the *de se* may achieve the modest goal described above, the more ambitious goal must remain beyond its reach.

2 The relativist account and its virtues

2.1 The relativist framework

2.1.1 The central idea

The central claim common to all relativist theories of *de se* thinking is that essential egocentricity is not a matter of self-reference. Quite the opposite: its distinctive epistemic profile is explained by the absence of self-reference. The content of thoughts whose ideal expression would be statements such as:

(1) I'm hungry.

(2) I'm feverish.

does *not* contain any constituent meant to designate the person whose mind harbours the thought in consideration. Further, *de se* thoughts are not even singular thoughts at all: they are *objectless*. Contrary to what the visible layout of linguistic representations such as (1) or (2) suggests, the intentional properties of the *de se* attitudes they ultimately give expression to are not to be identified by referring to classical propositions, that could be decomposed into a subject and a predicate[13]. They should be identified, more simply, by reference to the sole predicate (in the case of (1) and (2), the attributes "Hungry(x)" and "Feverish(x)", respectively). The information encoded by my mental state, when I entertain the thought expressible by sentence (1), thus takes the form of an impersonal content *à la* Lichtenberg, such as [[(1)]] or [[(1)]]'[14]:

(1) I'm hungry

13 Nothing essential hinges on this terminology, appropriate to a structural model of propositions; the relevant contrast can equally be drawn within a possible-world model of propositions, using the notion of "possible centred worlds" (Lewis 1979).

14 See for instance Lichtenberg (1971), ii, 412, §76.

[[(1)]] There is hunger

[[(1)]]' Hungry!

How could such an impersonal content account for the essential egocentricity of a *de se* attitude, such as the attitude expressed by an utterance of (1)? The relativists' answer is that the simplified, subjectless content they ascribe to (some) *de se* attitudes must be relativized, when it comes to truth-assessment, to a circumstance of evaluation that is richer than just a possible world. This richer circumstance of evaluation, or "index", can be formally described as an *n*-tuple that includes, besides a possible world *w*, several extra parameters: at a minimum, an individual of reference *s*, called the "agent" or "centre"; but also, very often, a time *t*, a location *l*, or even further parameters of relative evaluation (collectively designated below by the letter *k*), such as standards of taste or precision, moral or epistemic norms, etc.

Consider, for instance, the conditions that must be examined for the truth-assessment of the *de se* content a person *s* can express by saying

(3) It hurts!

For the relativist, this content is reduced to a mere predicate, approximately "Hurt(*x*)". The truth-making circumstance against which one should assess the truth-value of this simplified content (hereafter "centred content") isn't just the actual world $w_{@}$, – indeed, it wouldn't make much sense to ask whether "It hurts!" holds of the actual world – but a complex set of coordinates $<w_{@}, s, t, l, k>$. One only gets in a position to judge whether an assertion of (3) is true or false when one asks whether it is true that "there is pain" at time *t* and on the spot *l*, for the person *s*, relative to the standards of sensitivity to pain *k* of *s*.

In a word, the overall relativist strategy consists in impoverishing the encoded informational content, while commensurately enriching the circumstance of evaluation with new parameters.

2.1.2 The self takes leave

The main originality of the relativist account of *de se* thought, therefore, is the claim that it can dispense with the notion of an explicit representation of the self. Attitudes *de se*, while they realize the most primitive kind of self-knowledge, are paradoxically taken to be, as Perry (1986) provocatively says, "selfless".[15]

The underlying justification is that some essential aspects of the indexical mode of signifying, as soon as one moves from the realm of linguistic communication to that of mental representations, can be externalized, and in fact should be expected to be externalized, for reasons of cognitive economy. Some of the aspects of reality which play a semantically relevant role with respect to a thought do so, not in virtue of representational properties of this

15 See also Recanati (2007a), p. 176.

thought, but simply in virtue of the factual, empirical relations that the thought regularly stands in to its environment, thanks to being located among other material objects which together form a concrete context. The *intentional* relation postulated to exist between the self (as subject of a *de se* thought) and itself (as object of this very thought) within any self-reference theory of the *de se* can be replaced by merely *dynamic* relations between the self (as the causal source of a *de se* thought) and the environment witnessing the event of the thought, an environment that the identity of the thinker contributes to individuating. As Perry (1986) puts it: instead of being *about* the subject (an internal, intentional relation), attitudes *de se* only *concern* her (an external relation). It is true that the *theorist*, to be able to completely spell out the truth-conditions of a *de se* thought like "It's too hot", must specify the identity of the subject who entertains it. But the *subject* herself, at the level where she immanently grasps the present occurrence of her own thought, and as long as she doesn't take a reflective stance on it, doesn't have to think *of* herself.

2.1.3 Structural invariance as the source of person-relativity

What allows the cognitive economy that, according to relativists, distinguishes basic *de se* thoughts? Since none of the constituents of such a thought represents its author, in virtue of what does the thought "concern" her specifically, rather than any other person?

The reply given by Perry and Recanati rests on the notion of structural *invariance*. In Perry's terms:

> Sometimes all of the facts we deal with involving a certain *n*-ary relation involve the same object occupying one of the argument roles. In that case, we don't need to worry about that argument role; we don't need to keep track of its occupant, because it never changes. We can, so to speak, pack it into the relation. (…)[16]

In the case of *de se* thoughts, the subject constitutes just such a fixed object: none other than me, as a matter of fact, happens to entertain the relations to external objects and to my own body that my perceptions, my inner sensations and all my experiences reflect. This invariance places an architectural constraint on who is concerned by a given *de se* thought (i.e., who the "centre" is), as the source of these representations and the point of origin of the perspective from which they are contemplated. This structural constraint, in turn, prompts a cognitive economy on the part of the subject: it dispenses with an explicit representation of the self. Such an explicit specification would be idle, since the structural constraint just mentioned forces the choice of a value for the "subject-parameter" or centre.

In view of the foregoing, one of the main virtues of the relativist account of the *de se* is its parsimony. Not only is it good theoretical policy, in general, to refrain from postulating the existence of entities (representational entities, in this instance) that can be done without. But in the case at hand, the economy is also a cognitive one: it means that the subject, to be

16 Perry (1998), p. 4.

capable of *de se* thinking, doesn't have to harbour representations as complex as existing self-reference theories require them to be. Insofar as there is reason to think that less sophisticated, non-linguistic creatures can also have egocentric mental states,[17] this gives all the more psychological plausibility to the relativist option. Given this virtue, it would seem to fall on advocates of the competing referentialist accounts to provide an independent justification for introducing complex representations in their *explanans*.

2.2 The relativist explanation of the epistemic privileges

I will now introduce two recent examples of how the relativist theory has been applied to explain *de se* epistemology, illustrating what I called the "ambitious goal" in Section 1.2. These examples concern two of the epistemic privileges found in connection with *de se* thinking: the superior discriminative power of self-locating information, and immunity to error through misidentification.

2.2.1 Self-location

All *de se* thoughts go with a *quantitative* epistemic advantage, providing us with *more* knowledge than we would be able to gain in their absence. Intuitively, the *de se* mode of thinking reflects the fact that subjects represent reality in a *situated* way, i.e. from a particular *perspective*. But thinking from a subjective perspective encompasses simultaneously a certain limited, constrained awareness of the world, and a (more or less implicit[18]) awareness of the point of view from which the world is so presented. The reflexive capacity that *de se* thinking involves on the part of the subject is a grasp of her own presence in the world, at the centre of the perspective from which reality is given to her. That a *de se* thought tells me not just about the world but about my own location in it means that there is always a respect in which it is *more informative* than a perspective-neutral counterpart. Entertaining *de se* thoughts thus gives us access to richer information – what we may call "self-locating information", after Perry and Lewis – than we would have if we didn't represent reality in an egocentric framework, but only in an a-centric fashion, like an exhaustive encyclopaedia or an omniscient god would.

The irreducibility of self-locating knowledge to any amount of perspective-neutral knowledge is evidenced, *in absentia*, by a famous thought-experiment due to Lewis[19]:

> 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 at their world. (...) 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 top of the coldest mountain and throws down thunderbolts. Neither one

17 See for instance Proust (2007).

18 Exactly how far this implicitness can go is what is in dispute between self-referential and "selfless" (no-reference) approaches to the *de se*.

19 Lewis (1979), pp. 520-522.

> knows whether he lives on the tallest mountain or on the coldest mountain; nor whether he throws manna or thunderbolts.

Knowing all the propositions that are true at their world, the two gods possess objective omniscience. This, however, is intuitively compatible with a residue of ignorance as to their subjective position: each of the gods may still be unable to know which thing *he* is within that world, and to tell *himself* apart from the other god. The missing piece of knowledge is one that only a *de se* thought ("*I* am on the tallest mountain!") could yield.

2.2.2 Divine ignorance explained by fine-grainedness

To form *de se* attitudes is to put oneself in a position to gain new knowledge – precisely the kind of knowledge that Lewis's two gods lack. Lewis, and the other relativists, treat this contrast between *de se* attitudes and impersonal attitudes as a difference in their respective *fineness of grain*. *De se* attitudes, they claim, possess a special kind of content, structurally more simple since it is reduced to a predicative component, but also, and correlatively, more discriminating in principle than any non-indexical content.

This is brought out by the fact that the relativist semantics within which "selfless" *de se* thoughts are described has *more expressive power* than the standard propositional semantics. To entertain a classical, subject-predicate belief ("That *a* is F") is to depict the actual world as being a certain way. To have a *de se* belief is to do something more. Relativists account for *de se* attitudes by treating them as acts of implicit self-attribution of a special kind of property[20] (e.g. "being in pain", "standing on the left of Lady Liberty"), that Egan (2006a,b) calls "centring features" or "predicaments". These properties can fluctuate, within the same world and for the same individual, with the latter's changing perspective – roughly, the spatio-temporal point of insertion of her body in that world. The epistemic alternatives knocked down by such self-ascriptions are thus defined not simply *across* possible worlds, but *within* the one we actually occupy. *De se* thoughts, therefore, delineate a strictly larger set of *possibilia* than perspective-neutral, propositional thoughts.

The epistemic situation of Lewis's two gods is defined by a perfect knowledge of the world they inhabit, i.e. knowledge of all the classical *propositions* that are true at this world. But the finer modelling of content in terms of centred *properties* makes room for the residue of ignorance they display. They know exactly, of all the worldly *possibilia*, which one is instantiated by the world they belong to; but they don't know which individual *possibilium*, or predicament, they themselves embody inside this world.

Thus, the changes in a subject's cognitive state that *de se* attitudes are responsible for, and which allow for self-location, can be treated as *intentional* differences. Attitudes of this

20 The treatment of the objects of *de se* beliefs as *properties* attributed to individuals reflects Lewis's particular ontology, which excludes trans-world individuals. Leaving aside this controversial metaphysical background, the gist of Lewis's proposal can be recaptured within a standard modal logic framework by saying that, in a *de se* belief, a *centred-world proposition* is attributed to an agent. (As the distinction has no direct incidence on what follows, however, I will continue talking in terms of self-attributed properties, as contemporary followers of Lewis often do.) Thanks to Robert van Rooij for having pointed this out to me.

kind carry richer information than other attitudes do because their content is different, in a way that makes them true or false relative to something *more specific* than a possible world. As a result of this enrichment of the circumstance of evaluation, centred contents carve up the space of possibilities into a more precise grid than classical propositions do. What a *de se* thought represents thus imposes a stronger constraint on what should be the case to make it true. Hence the quantitative epistemic privilege attached to the possession of such thoughts.

I must express some unease with this solution. While it is true that the relativist semantics, *as used by the theorist*, has more expressive power, it is not clear in what sense her impoverished representations give the *subject* herself more information.[21] An explicit representation of herself as herself may well be redundant in her perspectival mental states, given certain psychological regularities; but while doing without such a self-representation may prove more efficient, it does not, by itself, enrich her epistemic standing. There is undeniably less data about the world to be read off a subjectless content, taken in isolation.

A possible reply on the part of the relativist would be to argue that describing the cognitive significance of *de se* thinking in terms of heightened discriminative powers and richer informational states might be just a manner of speaking, and a misleading one after all. Lewis, for instance, comments as follows on the situation of his two gods: "The trouble might perhaps be that *they have an equally perfect view of every part of their world*, and hence cannot identify the perspectives from which they view it.[22]" This suggests that the presumed extra richness of *de se* thoughts may just as well be described as the effect of a cognitive limitation. Being aware of my own location in the world, and of the fact that I view it from a specific point of view, presupposes that I do have such a point of view, i.e. a limited experience of reality. Egocentricity may primarily be a cognitive imperfection, even though this very limitation, because it is a structural constant, may be exploited for more economical cognitive design, and give rise in us to more kinds of representations than there would otherwise be occasion for.

I am not sure how convincing this move would be: it is hard not to think of self-location as a positive cognitive achievement, whose impairment translates directly in degraded performance[23]. I must postpone further discussion of this and related worries, however, to the last section. Let me now turn to the second example illustrating the epistemological agenda of (some) relativists.

2.2.3 Immunity to error through misidentification

De se thoughts come with epistemic privileges that are not only quantitative but also qualitative: the information they yield is not only richer, but also benefits, in some respects, from a higher degree of epistemic security. One of the forms that this epistemic security takes is a kind of certainty first described by Wittgenstein (1953), and referred to since Shoemaker

21 I thank Stephan Torre for questions that helped me become aware of this difficulty.
22 Lewis (1979), p 522. My emphasis.
23 See Perry's messy shopper (1979), and his library amnesiac (1977).

(1968) as "immunity to error through misidentification" (henceforth IEM). In intuitive terms, judgments that enjoy this kind of immunity are protected against errors having to do with the identification of the object they bear upon. They are impervious to the kind of mistake that would consist in attributing to the wrong object a property one knows to be instantiated. For instance, if I judge (on the basis of ordinary introspection) what I would express by declaring "I am hungry", this judgement is IEM: while I *could* arguably be wrong in some ways (i.e. by mistaking for hunger what is really a stomach ache), *one* way I *can't* be wrong is by misidentifying the person that my judgment applies to. If, based on interoceptive and introspective grounds, I am justified in thinking that someone is hungry, then I am also justified (on those same grounds) in thinking that *I* am hungry.

Restricting oneself to the case of *de se* thought, one can define $\text{IEM}_{de\ se}$ as the property that attaches to an occurrent *de se* judgment (and derivatively, to the person who makes it) if the following holds:

> **$\text{IEM}_{de\ se}$:** If I make, based on grounds G, a *de se* judgement of the form "I am F", then one way I can't be wrong is by knowing (based on G) that property F is instantiated, where the bearer of F is a person *s* that is distinct from me.

Some authors (e.g. Evans 1982, Pryor 1999) have conjectured that IEM judgments are based on a type of epistemic sources that simultaneously yield information about the instantiation of a property, and about the bearer of that instantiation. Inner sensation, for example, is such a source: the information it gives me about the presence of some bodily properties (hunger, pleasure, limb movement, etc.) is inseparable from the information that *I myself* have these properties. All the modalities of perceptual experience, as well as, to some degree, quasi-experiental epistemic attitudes such as episodic memory and situated imagination (Higginbotham 2003, Recanati 2007), share this peculiarity.

Of course, not all *de se* thoughts derive directly from experience; some are acquired through more detached, perspective-neutral epistemic sources, like testimony (my parents telling me the time of my birth) or inference (my belief that I am the winner of the election, based on calculation from the first partial results of the polls). Accordingly, not all *de se* thoughts are IEM[24]. Only those *de se* thoughts that are immediately based on certain kinds of experience in the right way, and hence reflect in a privileged manner the perspectival dimension of a subject's basic awareness of reality, are IEM. Recanati (2007) proposes to call all of the relevant *de se* thoughts "perspectival mental states". This subclass of *de se* thoughts are arguably the most primitive and fundamental ones, both in an evolutionary sense (experience plausibly comes before more abstract and a-centric modes of thinking) and from the point of view of their central role in justification (I will expand on this in Section 3.1

24 Conversely, not all thoughts that are IEM are, narrowly speaking, *de se* thoughts. It has often been noted, for instance, that demonstrative thoughts formed in the normal (experiential) way also exhibit IEM (Shoemaker 1968, Evans 1982, Wright 2012). This may also be true of temporal and spatial thoughts (McGinn 1983). As all these subtypes of indexical thinking deploy an egocentric framework of representation, however, there is still a clear sense in which IEM attaches primarily to the *de se*.

below). Experiential or "perspectival" *de se* thoughts will be the main focus in what follows.

2.2.4 The mystery of IEM dissolved by the absence of a representation of the self

According to relativists, it is the superior fineness of grain of the information carried by *de se* attitudes that explains the quantitative aspect of the epistemic privilege attached to them. As we'll now see, the qualitative aspect (IEM), in turn, can be accounted for by the feature from which this extra discriminative power stems, namely the absence of an explicit representation of oneself.

The relativist hypothesis is that no person at all is in fact represented in an egocentric thought like "It hurts" or "Hungry!". This entails two consequences. Firstly, at the level of *content*, no misidentification of the subject instantiating the property *F* presented in the thought can occur, simply because the general issue of identification is irrelevant here. Since they involve no representation of the self, basic egocentric thoughts require no effort of self-identification on the subject's part, which excludes *a fortiori* any risks of a *mis*identification.[25] Or more simply: $\text{IEM}_{de\ se}$, as defined above, is trivially true of basic *de se* thoughts, because the antecedent of the conditional, in their case, is false.

Second, at the level of *truth-evaluation*, when the identity of the person concerned by a *de se* thought is made explicit in the metalanguage (or in the subject's own reflective thoughts about that first thought), no misidentification is possible either. Recall that, for the relativists, the reason why there is no need for an explicit representation of the self at the level of content is that the identity of the subject who is concerned by this content is an *invariant* parameter of the context in which that subject thinks (or of her "perspective" on the world). Hence, when it comes to truth-assessment, the selection of the individual relative to which the content is to be evaluated is *automatic.*

3 Limits of the Relativist Account

The ingenuity and elegance of the way in which relativism accounts for the original epistemic profile of *de se* thinking mustn't cause one to underestimate the difficulties raised by the theory, which I will now consider.

3.1 Logical *vs. de facto* immunity to error through misidentification

To summarise the previous steps: in the relativist framework, the explanation of IEM has two components: *person-relativity* and *invariance*, the former deriving from the latter. The first component (person-relativity) is the absence of a representation of the subject in her own (basic) *de se* thoughts; or equivalently[26], the fact that these thoughts cannot be evaluated for

25 This can be seen as a radicalized re-interpretation of Evans's intuition that *de se* thoughts are "identification-free" (Evans 1982). Why "radicalized" is the topic of another paper (Guillot ms).

26 This equivalence rests on the assumption of what Recanati (2007a) calls the "distribution principle"; namely, the idea that if a parameter is part of the index, then it cannot be also represented in the content that is evaluated at this index. Ninan (2010) challenges this principle; but for reasons of space I will take it for

truth unless their truth-value is relativized to an index containing not just a possible world, but also (at least) an individual – the "centre". The second component is the invariance of the latter's identity, which is what makes it possible for her to do without an explicit mental designation of herself, excluding *a fortiori* the possibility of an incorrect designation.

Does this explanation, however, really capture the modality involved in the canonical formulation of the IEM principle? Arguably not. That I always occupy the epistemic perspective defined by the spatio-temporal location of my own body is only *de facto*; it could in principle be otherwise. This would be the case in a science-fiction scenario, in which my neural endings were hooked, via radio-transmission maybe, to the perceptual organs of another person.[27] Within such a scenario, I would still be able to think vicariously, "There is a light on the left" or "It's too hot"; but the light would be located on the left of the other person's body, and it is she who would be affected by the heat. In short, *she*, not I, would be *concerned* by my basic *de se* thoughts. The logical possibility of such cases of deviant causal chains shows that the relativist theory only accounts for a type of IEM that is *hypothetical*: the property that I detect to be instantiated, on the basis of a particular experience, cannot fail to be true relatively to me, *on condition that* the situation be normal, that my perceptual system work smoothly, that I not be hooked to the sensory organs of someone else, etc. The notion of IEM captured by the relativist theory brings into play, at best, a *nomological* form of necessity, dependent on the laws of nature in our world; or more plausibly, a mere contingent regularity, defeasible even within the actual world – it wouldn't take that much scientific progress, indeed, for the scenario of bodily transfer suggested above to become a realistic one.

But the notion of IEM that Wittgenstein or Shoemaker had in mind, and that an accurate description of mental phenomena arguably demands, is a stronger one. Shoemaker (1968)[28] makes a distinction between mere 'circumstantial' or *de facto* IEM, and 'absolute' or 'logical' IEM. Consider the contrast: given, on the one hand, the conceivability of neural transfer scenarios evoked in the previous paragraph, it makes sense to say that another person could have had my hunger or my fever, where these words refer to states of my body (causing certain sensations). The impossibility of an error as to the person involved, in this first, weak sense of IEM, hinges on the supposition that my phenomenology is normally connected to the states of this particular body, to the exclusion of any other; it is, therefore, a *circumstantial* or *conditional* form of impossibility. But on the other hand, it doesn't make sense to say that someone else could have had my *subjective experience* of hunger or feverishness. That an occurrent experience be given to my phenomenal consciousness is necessary and sufficient for this experience to be mine, and mine exclusively. Therefore, when the property that one self-ascribes, on the basis of introspection and/or somatosensation, is the *experience* itself, rather than the corporeal state that it reflects in normal circumstances, the kind of IEM attached to the predication isn't just hypothetical, but absolute. And the modal profile of this

granted here.

27 One could use the thought-experiment in Dennett (1978), with minor alterations. One important constraint is that the "host" must remain conscious throughout the experiment, so as to forestall any doubts as to whether *two* subjects are continuously present.

28 P. 557.

stronger form of IEM doesn't seem to be captured by the relativist theory.

This is an important limitation, for logical IEM is a ubiquitous characteristic of *de se* thought, and arguably its hallmark. As Shoemaker convincingly argues,[29] the stronger form of IEM is the more primitive one; any self-attribution brings it into play, either at its own level or at some level of the underlying justificational architecture. When I self-ascribe a state of hunger, for instance, even on the construal where this is merely a property of my body, it is always on the basis of some experience: the subjective experience of hunger, in the most ordinary case, or the experience of facts that are regularly correlated to the bodily state of hunger (e.g. the fact that the level of glucose measured in my blood goes below a certain threshold). The first self-attribution (of a corporeal state) presupposes in the background, in an explicit form or not, the more fundamental self-attribution of the experience, whatever it may be, that justifies it; and there, the stronger form of IEM is evinced. But if *any* self-ascription entails the presence of an absolute form of IEM in its grounds, then it is the latter phenomenon that a theory of the *de se* should target as its foremost *explanandum*.

At this point, a relativist could just accept the consequences of the theory, and argue that all IEM is really circumstantial. But she would owe us some explanation for why we sense a contradiction in the idea that what are phenomenally given to a subject as her experiences could in fact be someone else's.

Another option is to try and refine the theory so that it covers absolute IEM[30]. Here is what would seem to be the most promising way to do it. The relativist account rests on the notion of an invariance that makes explicit self-representation redundant. But one can distinguish two kinds of invariance, of distinct modal forces. The first invariant is that, whenever I am presented with an occurrent experience, *I* am the experiencer. The second is a regular correlation between what experiences *I* have and how things are in the world, relative to *me*. The mind is so constructed that the experiencer is normally also the centre (of perception and action) that the perspectival information it delivers is relevant for. The first invariance (the identity of the experiencer) is absolute, but the second (the coincidence between experiencer and sensorimotor centre) is only circumstantial – a mere matter of efficient cognitive design.

This places a *double* architectural constraint on the value of the parameter corresponding to the person concerned by *de se* thoughts (the "centre"). When all goes well, the two invariants point to the same individual. Both weak ("bodily") IEM and strong ("experiential") IEM are therefore satisfied. But when the centre of consciousness and the sensorimotor centre come apart, as in the science-fiction scenario above, only the first, stronger invariant remains in place. This account correctly predicts that, in this scenario, a minimal but absolute form of IEM (bearing on the ownership of the experiences themselves)

29 *Ibid*., p. 566.

30 Recanati does acknowledge the distinction between *de facto* and strong IEM, but doesn't explain the latter, which he seems to take as a primitive, *sui generis* property of consciousness. (See Recanati 2007, pp. 149-154; and his 2010, pp. 289-292.) He thus treats the two forms of immunity, not as two degrees of the same phenomenon, but as distinct in nature.

can survive the defeat of the more substantive but weaker form of IEM (bearing on the ownership of the properties represented by the experiences). And this strong form of IEM is explained by reinstating the now-familiar relativist move at a more fundamental level: as the identity of the experiencer never varies, there is no need for a representation of herself in her *de se* thoughts, hence no room for error as to this parameter.

This, however, leaves much to be desired. This extended "selflessness" account doesn't explain how I come to *recognise* my experiences as mine. The relativist describes *de se* thinking in a way that makes the satisfaction of the IEM principle trivial. This misses the deeper point about IEM. The issue isn't just the negative point that, in thinking *de se*, I'm not in a position to make mistakes of a certain sort – after all, a state of absolute ignorance, of radical suspension of belief or even that of a stone would also rule out such mistakes. The deeper problem that IEM theorists have sought to understand lies in a *positive* cognitive achievement: how is it that I get to the substantive (and infallible) knowledge that my experiential properties are *mine*?

This leads to a broader concern, orthogonal to the question whether the form of IEM captured by relativists exhibits the right modality.

3.2 Epistemic *vs* metaphysical immunity to error through misidentification

One might still consider that the relativist framework does provide an account of what I will call the *metaphysical* phenomenon of IEM, albeit with a weakened modal profile, corresponding at best to a nomological, rather than logical, form of impossibility of error. But what it leaves unexplained is the *epistemic* phenomenon of IEM.

Remember the informal description of IEM: if I think (based on an occurrent experience) a *de se* thought that is IEM, for instance "I am hungry", then one way it cannot possibly be wrong is by picking the wrong person, i.e. by being true of another person. This widespread formulation, as Dokic (2005) notes, is ambiguous between two modalities, an epistemic one and a metaphysical one:

> **Metaphysical IEM**: if a certain property (say hunger) is accessed by me through experience, then this particular instantiation of the property couldn't be true of another person.
>
> **Epistemic IEM**: if a certain property is accessed by me through experience, then (should I consider the issue) I couldn't fail to know that it's true of me.

While epistemic IEM arguably implies metaphysical IEM, the converse is not true. Epistemic IEM, therefore, is the stronger condition of the two.

According to the relativist theory, that (some) egocentric thoughts are IEM is explained by the fact that their truth-value should be assessed relative to the identity of the person who

fulfils the role of "centre", and by the *de facto* invariance of the latter parameter. The mental states I'd be inclined to express by saying "There is a light on the left" or "It's too hot" cannot fail to concern *me*, because I am the person who, in normal circumstances, invariably occupies the perspective from which perceptual and interoceptive information is conveyed to me. This reasoning does explain (*de facto*) *metaphysical* IEM; but it doesn't, on the other hand, shed any light on epistemic IEM. Since *I* am in fact the (unique) subject of a certain ego-implicit thought, whose purely predicative content is selfless, and which must be relativized to an architecturally designated person to be truth-evaluable, the information carried by that content can only concern the person I happen to be. But why is it that I can't fail to *know*, should I consider the issue, that this is indeed the case? That is what the relativist semantics, as such, doesn't explain.

Recanati (2007a) supplements the account by appealing to the contribution of what he calls "psychological modes". This expression designates the attitudinal component of an intentional state, i.e. the modality – cognitive, conative or affective – under which the encoded content is grasped (e.g. as a desire that *p*, a fancy that *p* or a belief that *p*). Recanati suggests that certain modes are special in that, whenever I grasp a centred content under any of these privileged modes, I know *ipso facto* that this content can only concern me. Thus, a content entertained under a kinesthetic mode, like one that would incline me to say "I'm about to lose my balance", or "I'm crossing my legs", informs me of properties of a body which is (in normal circumstances), my own. Similarly, my episodic memories, my exteroceptive perceptions and my internal sensations, although their content is devoid of an explicit representation of myself, always place me at the centre of the perspective from which information is delivered. To detect in my stream of consciousness the phenomenal modality associated with memory, exteroception or interoception thus naturally informs me that it is *me* who is concerned by the person-relative content currently crossing my mind.

This answer attributes an implausibly demanding role to the faculty of introspection, which enables us to discriminate different psychological modes. There are good reasons to think that our knowledge of the modality under which we entertain a given thought is far less secure than our knowledge of the identity of the person who is doing the thinking. Anyone can report having wondered whether the representation of a certain fact, vivid in one's mind, is the fruit of a dream from the previous night, a retrospective fabulation, or a genuine memory. In these cases, there is, however, no room for doubting that it is indeed *us* who are doing the thinking. But if access to the modality of the thought episode is far less stable than knowledge of the identity of its source, the former cannot be grounds for the latter. Recanati's relativist theory only replaces the original identification problem (how to identify the person at issue) by another (how to identify the type of the attitude instantiated), with no obvious explanatory gain.[31]

Notice that these concerns about epistemic IEM echo those raised in connection with the quantitative epistemic privilege. The account of the richer information carried by self-locating

31 See Wright (2012) for further objections.

attitudes seemed to trade on an ambiguity: while the relativist framework does give the *theorist* a finer grid to describe what is going on, the *creature*'s representations are in fact impoverished. It is true that centred contents, because they commit to less as to what reality should be like, do leave less room for error. But no *positive* epistemic gain is thereby achieved. In a structurally similar way, the relativist treatment of IEM does illuminate a way that the subject can't go wrong; but it doesn't explain how this creates certainty for her. In both cases, the theory's move is to lighten the representational burden from the subject's mind, and commensurately enrich the theoretician's toolbox. But this makes no less mysterious the psychological processes that support the subject's actual cognitive gains.

The relativist, of course, can reply that the impoverished representations she postulates are to be considered not in isolation but together with a larger environment or context, which makes them perspectival. This calls for two remarks. First, something should be said about how centred representations interact with that context, which would presumably involve low-level systems of detection of the relevant environmental features. But those extra independent assumptions, and not the relativist semantics *per se*, would then do the explanatory work. Second, as I will now argue, the context, to make centred thoughts perspectival, needs to include what was really the *explanandum*: a subject.

4 The limits of selflessness

The shortcomings of the relativist explanation of the characteristic epistemology that accompanies *de se* thoughts can be traced to a common source. The theory assumes a conception of subjectivity as the embedding of cognition in a certain perspective. This conception, as I'll try to show in the rest of this paper, is insufficient, at least if one uses the notion of perspective in the innocuous structural sense that relativists ostensibly give it.

4.1 Centre, perspective and subject

One of the attractions of the relativist model is that it takes seriously the fact that cognition, in certain respects, "adheres" to the context of its exertion, or is "anchored" in a local perspective. In its treatment of egocentric thinking, relativism takes as fundamental the notion of *perspective*. Basic *de se* thoughts, for relativists, are not representations of myself, but egocentric representations of other objects.

This much is quite plausible. But is the notion of a subjective perspective itself correctly analysed in the relativist framework? As I'll endeavour to show here, the relativist analysis leaves something to be desired, and an adequate approach should in fact go beyond mere person-relativity. The relativist way of capturing the perspectival character of *de se* thoughts rests on the introduction of centred worlds. Centred worlds are, in terms that Lewis (1979) borrows from Quine, "pairs of a world and of a designated space-time point therein."[32] Given a possible world, a "centre", according to this definition, is a simple *locus*, entirely

32 Lewis (1979), p. 531.

determined by the set of its coordinates in the dimensions of space and time. This, however, falls short of a perspective, in the sense relevant to describing the peculiarities of attitudes *de se*. A centre, as pictured by Lewis and his followers, is really only a point within space-time at which a subjective perspective may be anchored, not the subjective perspective itself.

For a distinction must be made between two importantly different senses of the expressions "perspective" and "point of view". The first sense is purely topographic, the second genuinely psychological[33]. Firstly, one can designate with these terms a location in the physical world. But while all concrete objects possess such a location, only a few – subjects – also possess what corresponds to the second sense of perspective, i.e. the faculty of exerting cognition from their point of view (in the first sense). Consider the intuitive contrast between, on the one hand, locating a panoramic viewpoint on a map, and, on the other hand, what it feels like to stand there as an observer and to have a look around. One cannot account for the viewing activity and the structure of the representations it originates only by invoking the location of the viewpoint. It is, however, this kind of reduction that would be required for the relativist explanatory project to work. By pairing possible worlds with the spatio-temporal coordinates of a person's body (the "designated individual") in the index, the centred-worlds framework does multiply the *possibilia* (the number of such pairs being strictly bigger than the number of worlds), and thus might explain – with the caveats we've seen – the richer informational import of *de se* thoughts. But it doesn't account for the other, qualitative dimension of the phenomenon of subjectivity. The centred-worlds model captures perspective as a *locus*, or as a potential observation standpoint; but not perspective as a property attached to the observer or to the act of observing.

The problem is that the notion of "centre" that would be needed for the theory to really offer what it promises is richer than that defined by Lewis. A simple point in the physical world isn't sufficient; what would be needed is a full-blown subjective consciousness, or more simply, a subject. Recanati (2007b) comes very close to acknowledging this point:

> (...) appealing to such [person-relative] propositions does not account for *de se* belief, because the person with respect to which such a proposition is evaluated can be thought of in many different ways or under many different guises. In particular, she can be thought of in a first-person way or in a third-person way. So what is distinctive about *de se* belief is not captured merely by appealing to person-relative propositions. This difficulty can be met, by introducing a new sort of situations for propositions to be relative to. Just as we distinguished persons from (other) objects, *we can distinguish subjects, or "first" persons, from other persons*.[34]

Subjecthood, in the classical approach, is defined as a metaphysical status that is bestowed on the basis of an epistemic achievement[35], namely the capacity to bear a certain

33 For a similar distinction, see Biro (1991, 2006) and Nagel (1974), p. 325. Vendler (1988), pp. 175-6, defends an opposing view, according to which any point of view in the psychological sense is, *in fine*, reducible to a purely topographical point of view, constrained by various structural requirements.

34 Recanati (2007b), pp. 145-6. My emphasis.

35 Of course, this classical (Cartesian, Lockean) notion of subjecthood as based on a mental relation to oneself is not the only one available. Some contemporary theories of the self have moved the emphasis from epistemic achievement to embodiment (Merleau-Ponty), agency (Anscombe, O'Brien), or commitment (Sartre, Moran). However, it is not obvious that those theories constitute alternatives as far as the point at

mental, non-accidentally reflexive relation to oneself, that most call "self-consciousness"[36]. It is, for instance, what Locke places in his definition of a person (a synonym, in this context, of what I call a "subject"):

> 9. Personal identity. This being premised, to find wherein personal identity consists, we must consider what person stands for; – which, I think, is a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places (...)[37]

Presupposing this non-accidental form of reflexivity as one of the basic ingredients of the theory, however, would raise several problems.

The first one is that positing a full-blown form of subjectivity as something that is already given in the context for *de se* thoughts to be relativized to would much reduce the economy, both theoretical and cognitive, that the relativist theory seemed to achieve in comparison to the competing self-referential theory.

Moreover, unless the relativist can show that subjects, as centres, can be constituted by something less costly than (some primitive form of) self-representation, it is not clear how she can account for the epistemology of the *de se* without postulating self-representations – or indeed, account for it by the *absence* of any self-representation. If positing subjects as centres turns out to be necessary to make the right predictions, then the move to substitute dynamic relations for intentional ones isn't entirely successful.

Another problem is that such a presupposition would seem to reduce the explanatory power of the theory. Subjectivity is a property that attaches to beings who have (or are capable of having) *de se* thoughts. Having the ability to form *de se* thoughts *just is* being a subject. Accordingly, one can legitimately ask from a theory of *de se* thinking that it takes subjectivity as one, if not the main one, of its *explananda*. Adopting it as one of the primitives of the theory, by substituting the epistemically demanding and metaphysically rich notion of *subject* for the more parsimonious Lewisian notion of a *centre*[38], would thus come close to begging the question: one would be taking for granted precisely what one was to explain. Still, it is such a primitive that the relativist theory must build on if it is to capture this defining peculiarity of *de se* thoughts that whoever originates one cannot fail to know, if she considers the question, that it is *she herself* who is the owner of this thought, and the centre of the experiential perspective it defines.

---

issue here is concerned. They all make it a necessary condition (although not a sufficient one) for subjecthood that the creature bear a special, mental relation to itself, e.g. bodily awareness or agent-awareness. This is asking for much more than a spatio-temporal location; which is my main point. Thanks to an anonymous reviewer for inviting me to clarify this.

36 I leave it open here whether or not this mental relation to oneself needs to be conceptually articulated in the subject's representations.

37 Locke (1694) II, XXVII, "Of Identity and Diversity".

38 Lewis himself compensates for the thinness of his notion of centre by another rich primitive in his theory, namely the notion of a *sui generis* attitude of implicit self-attribution, which replaces belief as the basic attitude, and through which the subject of a *de se* thought relates to the centred content she represents. This notion, which seems to play a crucial part in the modelling of *de se* thought, presumably presupposes full-blown subjectivity. But if subjectivity is already packed in the primitives of the theory, it is not something that the latter can *explain*.

4.2 A functional approach to subjectivity?

In the lines immediately following the passage quoted above, however, Recanati proceeds to downplay his concession that full-blown subjects are needed as centres, and offers a reply to the type of concern just raised:

> This is not an ontological move. I am not suggesting that we appeal to Cartesian *egos* in order to account for *de se* belief in the relativized-proposition framework. A subject is an ordinary person. What distinguishes him or her from other persons is only the contingent role he or she plays with respect to a tokening of the relevant relativized proposition.

The last sentence suggests a functional approach to "first-personhood" or subjectivity. The "contingent role" that the subject plays relative to a certain occurrence of a *de se* thought is an empirical relation, namely being the person who is causally responsible for the formation of the thought, and who is so situated as to be enabled to grasp its content by the mere fact of its occurring. Being so situated with respect to the thought has concrete effects on the other representations and on the actions of the person. Thus, to borrow a famous example from Perry (1979), if I fulfil the role of "first person" relative to a *de se* thought I would express by screaming "Bear attacking!" – the person, that is, who is "concerned" by the person-neutral content of such a thought –, this is manifested by a series of characteristic actions and attitudes (fearing, running for dear life, calling for help, etc.).

To sum up: in reply to the worry that a bare Lewisian notion of "centre" may not be enough to capture what we mean by a subjective perspective, Recanati offers a replacement notion that is still minimal enough not to beg the question as to what subjectivity, and *de se* thinking, consist in. In his view, the "subject-parameter", even if it is more than a geometrical point in space-time, needn't be interpreted in a metaphysically demanding way, as an entity constituted by a basic form of self-representation. It could just be identified with the system of the subject's body, states and actions, as functionally coördinated with her *de se* thoughts. This reply calls for a remark, and opens up a new difficulty.

First, a remark on the actual import of the concession. A functional definition of the "centre" or "subject-parameter'' is already an important amendment to the relativist theory. The first lesson to be drawn from this functional move is that the phenomenon of the *de se* is *not* reducible, as some relativists would hope, to the relativity of the truth-value of certain thoughts to the identity of the person thinking them. What makes a *de se* thought special cannot be captured by considering it and its distinctive semantics in isolation. What needs to be identified isn't just the person it "concerns", but also the complex set of further attitudes, epistemic adjustments and behaviours that surround its occurrence in this person's mind. To account for the peculiar epistemology of *de se* thinking, one needs at least to complete the relativist semantics with a certain kind of functionalist theory of mind, thus aiming at a mixed account.

But second, the functionalist response opens up a new problem. The revised, mixed relativist-functionalist theory defines *de se* thoughts as those representational states whose truth-value must be relativized to an index that includes at least a subject, defined in the functional way sketched above. But this seems to entail that any informational system that is endowed, as a matter of functional architecture, with detecting devices that are self-directed, and that is capable of coördinated reactions, would count as having (basic) *de se* thoughts and (a primitive form of) subjectivity. Consider, for example, the case of an electric iron with a thermostat designed to be sensitive to its own heat, and to prompt a reduction or increase of power whenever the system reaches, or falls below, a certain temperature. Does the lighting of the little lamp that signals "Not hot enough yet" amount to a basic *de se* thought? Intuitively, no. Yet the description of the electric iron's functioning fits the criteria that the relativist theory, revised as above, fixes for *de se* thinking. The functional approach to "centres" thus yields counter-intuitive predictions, and appears to be too liberal.

It is conceivable that a much more complex functionalist description could be proposed, that would suitably restrict the class of systems that qualify as subjects. But this is no small task, and it falls on the proponent of the selflessness theory to show how it can be done, as this is where the account's main *explanans* turns out to lie. Before this is done, it is not clear whether anything short of a full-blown subject will suffice in replacement for the original notion of "centre". For the time being, the theory seems faced with a dilemma, with the threat of trivializing the phenomenon of the *de se* on the one hand, and the risk of becoming question-begging on the other hand. The ambiguity on which the notion of "centre" trades, as long as the dilemma remains unresolved, points to a diagnosis of both the theory's initial appeal, and of the persisting dissatisfaction it inspires.

5 Conclusion

How far has the relativist got in her attempt to account for epistemic features of *de se* thoughts? Halfway at best, if what precedes is correct. As we saw in Section 2, the relativist framework can perhaps yield, with the reservation I mentioned, a convincing account of the quantitative epistemic privilege attached to *de se* attitudes, namely their fine-grainedness. However, Section 3 showed that the qualitative privilege (IEM) is only partially explained. Logical IEM (Section 3.1) and epistemic IEM (Section 3.2) are still unaccounted for.

I also argued that, if relativism is to make the right predictions, it must presuppose the existence of subjects as centres. Thus, as we saw at the end of Section 4.1, it implicitly treats the essential egocentricity of the *de se* as a *primitive*, rather than an *explanandum*. In other terms, it does not really reduce subjectivity to something more simple. Hence, the – undoubtedly precious, but limited – contribution of existing relativist accounts is to provide descriptive, rather than explanatory, theories of *de se* thinking.

A tentative conclusion could be that essential egocentricity – which includes the epistemic intimacy with one's own thoughts, the essential 'mineness' of experience, that the

property of absolute, epistemic IEM turns on – does not boil down to the limited nature of situated information. Relativist semantics captures the latter, but not the former. What it describes is a situation of partial information on the world: a much more common condition, surely, than that of being a subject. One may wish to restrict the use of the term '*de se*' to such limited access to the world, and to take a theory of the *de se* to be one that aims at modelling the semantics of partial information. It is unlikely, however, that this is what drew the attention of the generations of philosophers who have studied the 'first-personness' of experience. If this paper had only succeeded in making clear the distinction between those two projects – only the more modest of which lies within the reach of the relativist strategy – it would have achieved its goal.

Acknowledgements

Work on this article has been supported by funding from the European Community's Seventh Framework Programme FP7/2007-2013 under ERC grant agreement n°229441–CCC, during a research fellowship at CNRS.

I would like to thank audiences at various conferences and seminars, as well as Michael Murez, Stephan Torre, and especially François Recanati, for discussions that have helped me shape the ideas presented in this article. I am very grateful to Manuel García-Carpintero, Robert van Rooij, Crispin Wright, and two anonymous reviewers, for insightful and detailed comments on the article at various earlier stages. Special thanks are due to Conor McHugh for his reading, both acute and supportive, of several successive drafts. Any remaining shortcomings and errors are my own.

References

Alston, W. (1971). "Varieties of Privileged Access." *American Philosophical Quarterly*, 8, 223-41.

Anscombe, Gertrude E. M. (1975). "The First Person." In Samuel D. Guttenplan (ed.), *Mind and Language*. Oxford : Clarendon Press, 45-65.

Biro, John I. (1991). "Consciousness and Subjectivity." In *Philosophical Issues*, Vol. 1, *Consciousness* (1991), 113-133.

Cassam, Quassim (ed) (1994). *Self-Knowledge*. Oxford: Oxford University Press.

Castañeda, Hector-Neri (1968). "On the Phenomeno-Logic of the I." *Proceedings of the XIVth International Congress for Philosophy* 3 (1968), 260-6. Reproduced in Castañeda, Hector-Neri, *The Phenomeno-Locic of the I*, ed. by G. Hart and T. Kapitan. Bloomington : Indiana University Press (1999).

Chisholm, Roderick M. (1981). *The First Person*. Minneapolis : University of Minnesota

Press.

DENNETT, Daniel C. (1978). "Where am I?" In D. DENNETT, *Brainstorms*. Cambridge, Mass. : The MIT Press.

DOKIC, Jérôme (2005). "Introspection, déploiement et simulation." In *Philosophiques*, Vol. 32, n°2 (fall 2005), 383-97.

EGAN, Andy (2006a). "Secondary Qualities and Self-Location." In *Philosophy and Phenomenological Research* 72:1, 97-119.

EGAN, Andy (2006b). "Appearance Properties?" In *Noûs*, 40:3, 495-521.

EVANS, Gareth (1982). *The Varieties of Reference*. Edited by John McDowell. Oxford: Clarendon Press.

FREGE, Gottlob (1918). "Der Gedanke. Eine logische Untersuchung." In *Beiträge zur Philosophie des Deutschen Idealismus* (1), 58-77.

GEACH, P. (1957). "On Beliefs About Oneself." *Analysis* 18, 23-24.

GERTLER, Brie (ed) (2003). *Privileged Access. Philosophical Accounts of Self-Knowledge.* Aldershot: Ashgate Publishing.

GUILLOT, Marie (ms). "Identification-free at last: Evans, semantic relativism and immunity to error through misidentification."

HIGGINBOTHAM, Jim (2003). "Remembering, Imagining, and the First Person." In A. BARBER (ed.), *Epistemology of Language*. Oxford : Oxford University Press, 496-533.

LEWIS, David (1979). "Attitudes *De Dicto* and *De Se*." In *The Philosophical Review* 88, 513-43.

LICHTENBERG, G.C. (1971). *Schriften und Briefe.* Carl Hanser Verlag.

LOAR, B. (1976). "The Semantics of Singular Terms." *Philosophical Studies* 30, 353-77.

LOCKE, John (1690). *An Essay Concerning Human Understanding*.

MCGINN, Colin (1983). *The Subjective View. Secondary Qualities and Indexical Thoughts*. Oxford: Clarendon Press.

MILLIKAN, Ruth G. (2004). *Varieties of Meaning*. Cambridge, Mass. : The MIT Press.

NAGEL, Thomas (1986). *The View From Nowhere*. Oxford : OUP.

NINAN, Dilip (2010). "Semantics and the Objects of Assertion." *Linguistics and Philosophy* 33 (5): 335-380.

NINAN, Dilip (2008). *Imagination, Content and the Self.* PhD thesis. Cambridge Mass. : Massachusetts Institute of Technology.

PERRY, John (1998). "Myself and I." In Marcelo Stamm (ed), *Philosophie in Synthetisher Absicht (A festschrift for Dieter Heinrich)*, Stuttgart: Klett-Cotta, 83-103.

PERRY, John (1986). "Thought Without Representation." *Supplementary Proceedings of the Aristotelian Society* 60, 137-52.

Perry, John (1979). “The Problem of the Essential Indexical.” *Noûs* 13, 3-21.

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

Proust, J. (2007). “Metacognition and Metarepresentation: Is a self-directed theory of the mind a precondition for metacognition?” *Synthese*, 159, 271-295.

Pryor, John (1999). “Immunity to Error Through Misidentification”. *Philosophical Topics* 26 (1/2):271-304.

Recanati, François (2012). “Immunity to error to error through misidentification: What it is and where it comes from.” In Prosser and Recanati (eds) (2012), referenced above.

Recanati, François (2010). “Le soi implicite.” *Revue de métaphysique et de morale*, n°4/2010.

Recanati, François (2007a). *Perspectival Thought. A Plea for (Moderate) Relativism*. Oxford : OUP.

Recanati, François (2007b). “Relativized Propositions.” In Michael O’Rourke and Corey Washington (eds), *Situating Semantics. Essays on the Philosophy of John Perry.* Cambridge, Mass.: The MIT Press.

Shoemaker, Sydney (1968). “Self-Reference and Self-Awareness.” *Journal of Philosophy* 65 (19), Oct. 1968, 555-67.

Sosa, E. (1983). “Consciousness of the Self and of the Present.” In J. E. Tomberlin (ed), *Agent, Language and the Structure of the World.* Hackett, 131-43.

Stalnaker, Robert (2008). *Our Knowledge of the Internal World*. Oxford: OUP.

Stephenson, Tamina (2010). “Control in Centred Worlds”. *Journal of Semantics 27*.

Stephenson, Tamina (2007b). “Judge Dependence, Epistemic Modals, and Predicates of Personal Taste.” *Linguistics and Philosophy* 30;4.

Stephenson, Tamina (2007a). *Towards a Theory of Subjective Meaning*. Ph.D. Thesis, MIT.

Stojanovic, Isidora (2012). “The Problem of De Se Assertion”. *Erkenntnis* 76 (1): 49-58.

Torre, Stephan (2010). “Centered Assertion”. *Philosophical Studies* 150 (1): 97-114.

Vendler, Zeno (1988). “Changing Places.” In D. Seanor & N. Fotion (eds), *Hare and critics: Essays on Moral Thinking.* Oxford : Clarendon Press. 1988.

Wittgenstein, Ludwig (1953). *Philosophische Untersuchungen/Philosophical Investigations*. Edited by G.E.M. Anscombe et R. Rhees.

Wright, Crispin (2012). “Reflections on François Recanati’s, ‘Immunity to error through misidentification: what it is and where it comes from’”. In Prosser and Recanati (eds) (2012), referenced above.

Wright, Crispin, Smith, Barry C. and Macdonald, Cynthia (eds) (1998). *Knowing Our Own Minds*. Oxford: OUP.