RICERCHE

Reflections on Quasi-Indexicals, Self-Reference and Self-Knowledge[1]

Giuseppe Varnier[a]

█ **Abstract** Building on recent linguistic and philosophical research on quasi-indexicals, self-consciousness, anaphora, and *discours indirect libre*, I argue that they raise problems (as already pointed out by Castañeda and others) for the definition of (*de* se) self-knowledge understood according to the Classical Definition of Knowledge (conceived as expressing at least the necessary conditions for knowledge, if not a sufficient set of conditions). I call this extremely difficult problem the "non-detachment problem". I show that, for this reason, self-knowledge must always be considered perspectival and non-third-personal, in the relevant cases. I also discuss and criticize the Lewis-Chierchia interpretation of *de se* attitudes. Furthermore, I discuss the role of the (a) *self* in the strict sense in an adequate account of so-called *self-knowledge*, and in a problem closely related to the "non-detachment problem", that of the reconstruction of the *cogito*.

KEYWORDS: Indexicals; Quasi-indexicals; Classical (tripartite) Definition of Knowledge; Self-knowledge; Self-consciousness

█ **Riassunto** *Considerazioni su quasi-indicali, autoriferimento e conoscenza di sé* – Sulla base di recenti ricerche linguistiche e filosofiche sui quasi-indicali, l'autocoscienza, l'anafora e il *discours indirect libre*, intendo sostenere che essi sollevano difficoltà, già anticipate da Castañeda e altri, per la definizione della conoscenza di sé (nel senso del *de se*), se essa è pensata secondo la definizione classica di conoscenza considerata almeno come condizione necessaria della conoscenza. Chiamo questo problema di difficilissima soluzione il "problema della non distaccabilità". Cercherò di mostrare che, di conseguenza, la conoscenza di sé deve essere intesa come sempre "prospettica" e non di terza persona, nei casi rilevanti. Intendo poi discutere criticamente l'interpretazione di Lewis e Chierchia delle attitudini *de se*. Inoltre, discuto il ruolo del (o di un) *sé* in senso stretto in un resoconto adeguato della cosiddetta *conoscenza di sé* e in un problema prossimo a quello della "non distaccabilità": il problema della ricostruzione del *cogito*.

PAROLE CHIAVE: Indicali; Quasi-indicali; Definizione classica (tripartita) della conoscenza; Conoscenza di sé; autocoscienza

✠

THIS PAPER STARTS WITH THE well-known problem of non-substitutivity *salva veritate* (or better *salva propositione*) of pronouns in embedded quasi-indexical contexts (such as "John knows that he (himself) is sad") as originally highlighted by Castañeda with his famous John the Amnesiac example, and by many other examples in the literature.

[a]Dipartimento di Scienze Sociali, Politiche e Cognitive, Università degli Studi di Siena, Complesso San Niccolò, via Roma, 56 - 53100 Siena (I)

E-mail: giuseppe.varnier@unisi.it (✉)

But the paper focusses on cognate examples (in *discours indirect libre* and subjunctive moods) with a *factive* verb (the epistemic operator "know"), because (as I show in section 1) these allow us to describe what I call the basic problem, or the non-detachment problem, that is the failure of the rule "aKp → p", *and* they indicate some consequences of this failure both for the classical (tripartite) definition of knowledge, and, on the other hand, for the general problem of self-consciousness. In section 2, I draw on this point in order to discuss the status of so-called knowledge about oneself or self-knowledge.

I claim that self-knowledge enjoys a very particular status, which sets it apart from knowledge in the usual sense, and also in the sense of knowledge about a self. I suggest that alleged evidence to the contrary may well be derived from a linguistic and cultural bias. In section 3, the defense thus presented of the logical and metaphysical (*sui generis*) reality of private self-knowledge enables us to propose a criticism of the property-ascription approach to *de se* attitudes (and to propositional attitudes in general) advanced by D.K. Lewis and G. Chierchia. We show that this "solution" threatens to destroy the pre-linguistic idea of acquaintanceship with oneself (in the first person), which our analysis has established on the basis of both intuition and the impossibility of effectively analyzing self-knowledge on a par with all other kinds of factual knowledge. I defend the view that first-person attitudes at least are better understood as depending on limited-access propositions containing phenomenal, qualitative elements.

In the Conclusion, I sum up all previous arguments, showing how my thesis has consequences both for the understanding of the *cogito* argument and for the refutation of radical (reductive) materialism.

▪ The basic problem (non-detachment problem) and the classical definition of knowledge (CDK)

Let us start with:

(1) John knows that he (himself) is loved by everyone[2]

This sentence presents various problems, which I will explain in due time. If the reader thinks that such problems make it too difficult to handle, she may consider that, firstly, "everyone" means "everyone in his own circle" here. Secondly, she may consider that most of what I will have to say applies to a clearer *de se* variant such as (1'): "John knows that he (himself) is feeling sad now".

Now, there are in principle two possibilities. We can say that either (A), "John" and "he" have the same meaning, or, (B) "John" and "he" have different meanings – but the latter seems to be excluded by the specification "(himself)". I try to use "meaning" (no doubt a dubious translation) in Frege's original sense for *Bedeutung*. So it seems obvious that "John" and "he" have different *Sinne* ("senses"). As for meaning, we shall see that, even under the necessary assumption that merely linguistically the two expressions are coreferential, to go on and claim that this means that "John" and "he" have the same meaning may turn out to be problematic. In other words, we are interested here in the case that we can express as follows: it is of himself, John, that John knows that he is loved by all – i.e. linguistic coreferentiality. This is now easy to understand, so expressed, but it is a way of repeating the naive formulation of the very problem that should be solved. We can only start from here.

(1) provides a good example of a quasi-indexical sentence embedded in a cognitive (epistemic) prefix, or main clause. We take for granted the *linguistic* coreferentiality (though they are not synonyms!) of "John" and "he (, himself)", according to what is generally termed the (intended) *de se* interpretation in the Lewis (and Castañeda) tradition. *De se* is not *de re*; and if it can be conceived of as *de dicto*, this requires a lot of explaining as made clear, among other things, by the case of many languages, such as Spanish or Italian, which resort to infinitive clauses, where no pronom-

inal *dictum* at all is present (expressed) on the surface ("John knows that he, himself is loved by all" = in Italian "John sa di essere amato da tutti"). We use, in any case, a weak, linguistic concept of coreferentiality. We also assume that "he, himself" is in some sense more than a simple anaphora.

But then, in the subordinate clause of the sentence we started from, (1), the meaning (sense) of "he(, himself)" is not exactly that of the pronoun "he" (indexically or otherwise interpreted) in any usual *oratio recta* context – nor does it mean John, as is obvious. To see this, suppose that in "John knows that he is loved by all", "John" and "he" (say David) do not corefer. Then it is obvious both that "he is loved by all" follows (it can be detached, as I shall say, as a self-standing sentence, by factivity, that is truth-entailment) and that some name for the person designated by "he" ("David" here) could be (made) publicly available.

In the example that we presented, we could assume that no such name is available for John any more, not from his point of view: this is the essence of Castañeda's examples, such as John the amnesiac. Still he can think about *himself*, and this can be registered and understood by the ascriber (as we are doing right now). These are some of the main reasons why Castañeda introduced the notation "he*" for this particular usage of "he(, himself)" – as "quasi-indexical" – since it is linked to an antecedent by anaphora plus an ascription of *de se* attitude. We can make this even clearer.

In coreferential contexts such as the ones we are examining, "essentially or intrinsically anaphoric", as we said, means "essentially and in itself in *oratio obliqua*". One must add that this is particularly evident in the third person. In the first person singular (though there are quasi-indexical, with interesting properties) there is no problem achieving anaphoric coreferentiality, which is given automatically at every occurrence of "I". When we ascribe a discourse in the first person singular, we often profit from this continuously

anaphoric character of "s/he", as it is transferred from the first person singular.

There are special cases in which this continuous character is typically to be interpreted as anaphora of "he (, himself)" – or "s/he*" – in particular. One is *discours indirect libre*, in which there are no subordinate or that clauses (it would be ungrammatical to use "that"), but the effect reached is closely akin to that of quasi-indexicality, including *de facto* non-detachability, also because of that continuous character. Consider:

(2) John opened the door of the old house. He had returned home. He felt strange...

If this is correctly interpreted as *discours indirect libre*, or some equivalent, there is no doubt that what is recorded is what John feels and thinks, *from his own point of view*: "he had returned home" means here that he thinks that he, himself – and not John, or anyone else seen from a third-person perspective – has returned home (it would be completely compatible with amnesia). A certain use of subjunctive in German (and other languages) is also comparable in its induced effects. Consider:

(3) John sprach vor seinen Kollegen. Er sei dankbar. Er denke, daß ... etc.

Here it is clear that what is reported is the fact that John publicly *declared* that he was grateful, and that he *said* that he thought that..., etc. This is also connected to the essentially anaphoric (but not exclusively anaphoric) nature of quasi-indexicality. The point that we do not meet such-clauses in these usages is very important, because it confirms that the linguistic (and epistemological) nature of such expressions manifests itself in the following way. There are some truths which cannot really be separated from John's point of view, even though, within certain limits, they may seem to be expressible in main clauses (think of the German example above).

They cannot be obtained, or detached, from the knowledge or justified true belief John has of them, by factivity and perhaps even by MPP. In other words, the rule "aKp→p" fails in the context of such quasi-indexical examples, for the simple fact that there is no *oratio recta* sentence (and proposition, possibly) at all, that can be clearly obtained by the rule.[3] This is what I call the general detachment problem, which has already found its way into the literature. (Whether in addition there is a real, general threat for MPP is a topic for another paper

We just assume that the classical definition of knowledge expresses at least necessary conditions of knowledge, but, for simplicity's sake, we shall speak of knowledge simple. We write

CDK:
aKp =def
(1) aBp & (a believes that p)
(2) aJp & (a is justified in believing that p)
(3) p, (p, or p è is true)

If a q were a truth about oneself available to oneself in (quasi-)indexical form (in the way I described, but in the first person singular), then, for subject a, it must be true that aKq, assuming that there is a good, first personal (incorrigible) justification, and we know about it. This (justified true belief, hence knowledge, *according to the classical definition*) seems to apply well to sentences such as "John knows that he (himself) feels sad", which seem to be true, if sincerely entertained and/or expressed, by a) first-person authority, and, b), (Cartesian) privileged access. But, then, that aKq is true must – by condition (3) – imply that q, or that q is true. We could, from the truth of "aKq" and "aKq→q", derive "q" by modus ponens. But this is absurd, and we cannot do this. Even in the first person singular, if the sentence is thus embedded, it is easy to see that this sentence cannot in principle identify a proposition or sentence q independently from its essentially *oratio obliqua* (complete) expres-

sion; q is not a sentence in the sense of being (expressible in) a main clause. It makes sense only within a cognitive prefix, even one such as "(I know/feel that) I feel sad".

Similar problems affect the expressibility of the truth of q, that is the very meaningfulness of T(q). Yet we say something comprehensible when we say that John knows something when he knows (self-consciously or not) that he feels sad, i.e. that he feels sad. We can even distinguish between experiences that are simply conscious (non reflective), like sadness, and others that seem to require a degree of self-consciousness, like shame.

What we denoted, with approximation, by "aKq" seems to be, however, a form of knowledge. We cannot express exactly what it is knowledge of, though it entails a putative fact (that John feels sad, *from his own point of view*, and this is, notice, the only point of view that ensures sufficient justification – if he were judging that he, John, is sad by looking at himself in a mirror, that would not do). Knowledge comes before truth here, and even before justification, knowledge being defined by principles a) and b) in such cases. Epistemologically and metaphysically, it *is* knowledge. It looks as if it is not knowledge logically, as it does not comply with CDK.

We have a dilemma. (C): if there is knowledge that q, either (1) we have a counterexample to CDK, and the whole argument stops here, as we saw above; or, (2) q is simply not expressible autonomously. On the other hand, (D), we can rule that the epistemic subject a's simply (justified) true beliefs about himself from a first person perspective cannot be expressed except in *oratio obliqua* under certain constraints: they are not expressible as far as they are not detachable. In (C1) we have a problem for the general expression of the concept of knowledge. More than this, John's beliefs about himself in the first person are true, and a kind of knowledge, because they are justified and accessible thanks to a) and b), whereas the truth, condition ((3) of CDK) cannot even be expressed. So in case (C1) not only the logical

expression of knowledge becomes very difficult, but it becomes meaningless to speak about the factivity of *know* and such. It seems that there are severe limits to the validity and sense of CDK.

In (C2) and in the second case, (D), we have knowledge or truths that are, as complete, self-standing propositions, inexpressible, again a great difficulty for any logical definition of knowledge and truth, and a difficulty posed again by the non-detachment problem. This is so because "he (‚himself)" in *oratio obliqua* does not have the same sense as "John" in "John knows that John feels sad" (either in the main or the subordinate clause). It is clear that we can "detach" both "John is loved by all" and "he is loved by all" (with "he" a "normal" pronoun, or even indexical) when the reading is not *de se.*

The point is that these complete sentences do not express what John knows when he feels sad, or knows that he is loved by all, and which he can know without knowing that he, himself is (called) John. He does not know it about someone he conceives of through "he" simply, even if it should turn out that it is John himself (think again of mirror examples).

It is already clear that all these problems are close to the question of self-consciousness (and crucial for a thorough understanding of what self-knowledge amounts to, both in the sense of consciousness of one's "self", and, in a weaker sense, of consciousness by which we have true and/or justified beliefs about ourselves, or about some aspects of ourselves at least). In this section, I tried to highlight the main logical problem in a clear manner. I shall develop the theme of self-consciousness, and especially self-knowledge, in the next section and in the conclusions.

Some consequences for self-knowledge

Some historical remarks. It has been doubted, since Kant, that self-consciousness can give us knowledge in the sense claimed by Descartes. Yet we admit it can give us true and indubitable beliefs: this includes, it seems at first, such beliefs as one being oneself, or s/he himself or herself. So there has to be a fundamental distinction in self-consciousness between justified and unjustified true beliefs (about oneself too).

With these provisos, I speak about knowledge about oneself, though the terminological choice should be better addressed. In any case, the notorious distinction between justified true belief and knowledge proper should play no role in what follows, at least not a role that can be discussed in the present paper.

So through self-consciousness I can reach certainty – justified belief, accepted by others as *ipso facto* true – that, for instance I am thinking now, or I am sad now, or at least that I exist as thought (the *cogito*), or that there are thoughts (Lichtenberg's anonymous "es denkt"), or that I have a toothache (or other such predicates are instantiated). Even admitting that I could be wrong, these beliefs seem to be justified from a first-personal point of view, due to the subject's authority. And yet self-consciousness has intrinsic limits: it and introspection will not tell me how tall I am, or what colour my eyes are. This is not knowledge based on first person authority, even if it becomes deeply entrenched among our information about ourselves.

Non-physical predicates pose a difficult problem. Knowing that one is loved by all (all persons relevant to him or her) is probably a borderline case, and it seems unlikely that this can be determined by the subject without external help, *unless it is considered to be an introspective feeling or sensation*, which is also (barely) possible. (This is the reason why I chose my main example about John.) But knowing that one has an objective dental pathology (not just toothache), that one is moral, or that one is not lucky, all seem even harder to consider part of what is given within the limits of self-consciousness. In a similar spirit, Wittgenstein said that love for others is not a sentiment, for it must be *proved.* Certainly, self-knowledge has limits, and part of it must come from empirical or intersubjective

sources, not from within. Self-identification is necessary to such forms of self-knowledge. But this poses problems, to begin with, about the knowledge we can actually reach by way of self-identification, which is an open, never-ending process, in a sense.

Let us take stock. We cannot make sense of all this unless we make some implausible assumptions, such as that some of these truths are not truths about oneself in any way. But we excluded the possibility that there are no real truths about oneself. And this hypothesis is not so different from the claim that some such truths are not really truths about oneself in any case.

This is not so different from claiming that they are not truths at all. What is certain is that they are not "truths" that are easily detachable ones, at least *prima facie*. For different reasons, conditions (2) and (3) of CDK seem not to apply. If CDK is at least a "necessary criterion" of knowledge, it follows from our logical analysis that self-knowledge, which *cannot* be unpacked into the three conditions, is not a candidate for knowledge proper. (We would have reached the same conclusion if we had claimed that it is justified true belief, but not knowledge.) It is not even to be understood as permitting assertion about/ascription of knowledge (except primitively, to oneself, perhaps).

If knowability follows logic, and not metaphysics, Richard Moran is right, independently of his theses about self-knowledge and moral agency (lack of space prevents me from focussing on this), in stressing that it is not a "purely epistemic" gain. [4] If knowability is metaphysical, proper, *de se* self-knowledge remains fragile. One does not see how it can be about evidence, except in an empirical sense that involves experience but excludes knowledge of the independent existence of a self, a subject of independent predication. (Notice that we say: I am myself – but, if at all: I *have* a self.)

A surprisingly Kantian conclusion. Self-knowledge is there, but it is *essentially* first-personal and perspectival and close to non-

sharable, opening up a new dimension at best. (On the other hand, new acts of self-identification are conceivable that lead us to wider self-knowledge in this sense (e.g. when I add a new property I discover to myself, so to speak). Add that we can distinguish three orthogonal axes in this: (i) immunity from error through identification; (ii) first-person authority; (iii) incorrigibility. (i) is also present for the outcomes of acts of enlarged self-identification, and to some extent (ii) is also valid for resulting beliefs.

It might be objected even to my undeterred defence of the autonomous (to say the least) status of self-thinking, and self-knowing, that, even in the case of knowledge proper, we know a lot of things about what is known or wanted by other people about themselves. And that in fact we proclaim such knowledge by obvious attributions, also in the third person. As it can be objected in return, that the impression that we correctly do so may well be dependent on a linguistic bias (some would say, a cultural one).

In some languages, first person and third-person inner experience or volition attributions are not on a par. In a well-thought Japanese grammar (but the facts are known to any student of Japanese), we read: «Because of the way Japanese works, and the way the world is interpreted and thought about in the Japanese mindset, one never presumes to truly know what's going on in someone else's head. Because of this, you cannot say that "Bob wants an apple", because even though he might give off all the signals that he does, and even though he may have said so himself, you might still be interpreting the signals wrong, and he might have only said he wanted one instead of really wanting one. Because of this, rather than using [the straightforward desiderative form] for second/third-person desideratives, the classical helper verb [...] is used».[5] Such phenomena, related to the problem of evidentiality, are found in various languages. [6]

Forget the Author's quasi-Sapir-Whorfian reference to a "Japanese mindset", and his

likely but inessential conjecture that a certain Bob might have lied out of courtesy *saying* he wants an apple. There is no reason to assume that the Japanese see a different world, even less to suppose they are professional social paranoids (unlike Western people!). It stands to reason that, in such languages, it happens that the commonsensical evidence in favour of a mildly solipsistic stance with reference to third-person thoughts and attributions is lexicalized and grammaticized, whereas the same evidence is passed over in silence by, notably, the surface syntactic structure of English – with philosophical consequences for our theme. (Think what Descartes would have done, had he mastered Japanese.) Sure, linguistic quirks are no proof against common sense intuition (though this is not just a quirk), neither one way nor another: this is exactly my point.

More decisively, we cannot express autonomously what one knows exclusively about oneself in such cases. But *there is* something, and we cannot decide that all contents here are third-personal.

A criticism of D.K. Lewis' approach

David K. Lewis, in the article cited above (cf. *Postscript*), had an exhaustive idea to solve such problems. From the standpoint of linguistics, it works rather well, but leaves philosophical questions open. The typically *de se* "he (himself)" in the example – as developed by Chierchia (paper already cited, *passim*) – is accordingly read as a locally bound variable, one bound through a lambda – that is a predicative abstraction – operator. (There is a problem here. The idea of binding a variable in logic and that of binding in Chomsky are not obviously the same.)

The subject in the main clause is similarly bound. In short, this means that John enjoys the same property enjoyed by all those who know that they are loved by all, the property of knowing themselves (each one) to be loved by all. One problem is that each one is, in such cases, usually, loved, or may be loved, by

everyone *else except for himself or herself*, but let us put this aside. Complete notation apart, it is clear that this is not satisfactory for explicating *de se* thinking as self-access, but let us look at the solution closer up. The subordinate pronominal subject (either expressed or not) becomes a locally bound variable. The subject is not a constituent of a proposition, but self-attributes the property thus identified, and the property is also objective, since "know" is factive.

This is a solution for *de re* attitudes in general, but it is claimed it is useful especially for *de se* ones. I have doubts. Chierchia used it for the disturbing cases of *de se*, where avoiding propositions altogether makes things so much smoother. The subject of the main clause no longer entertains a proposition, but attributes a property to herself. But as we saw, firstly we are not told what the property means in itself, nor how it is assigned (that I am loved by all does not intuitively entail that I must be loved by myself). It is just generalized. This is perhaps inevitable. Secondly, however, and in close connection the difference between John attributing the property to himself and John attributing it to John is erased, but for the presence of John's centered point of view in the former case. Even if these problems were solved, the non-detachment problem would stay with us. What does John actually know, or feel, rather than just see in a way that can be captured in a purely linguistic manner, and is a linguistic convention?

For it seems indeed that the property that he attributes to himself, if it is that, does have a meaning that goes beyond abstraction from all similar cases (simply linguistically described), and *points to an inner experience*. This is true even if it is correctly foreseen that there is no proposition that can be deduced and "detached". Self-knowledge seems to disappear, and so does any entitlement to reflect (on) oneself. There must be something John knows – i.e. is true too. It may be difficult to see how this is encoded in a property, esp. in the epistemic case, with factivity. It is not said what they have in common, a

proposition *and* a property.

In fact, in principle we do not even have the certainty that all those who know of themselves what John knows of himself have the *same* experience. *This* is only reflected by the non-detachment problem. Even if we are not realists about *qualia* etc., there is here – given the language of fiction and *discours indirect libre* – also on a purely linguistic level, something that is objectively lost to us. It is the informational part of the first personal content from a proper, self-owned point of view. To see this is simple.

From my experience that (I know that) I am loved by all, if we accept it on first person authority, and incorrigibility, we can believe that the person I am is loved by all, *only* if we accept his or her word for it, or if we admit that his or her inner sensations or thoughts are translatable into ours. There is really no reason to do so, and the problem is much like the one about colour sensations, if we take first person authority, and privileged access, seriously. There is no doubt that I, or John, can know it. But what do we know exactly? Even in fully intersubjective language, "to be loved by all" is ambiguous, and the meaning of "know" often changes contextually. Can we say that John is wrong in believing what he believes, or *we* are, in acknowledging that what he believes *is* knowledge? Refusing the idea of propositions we are putting forth leads to such questions, and ultimately to ineffability.

The idea of an incomplete proposition is, at this point, in itself no more meaningful. Distinguishing between semantic and cognitive propositions does not solve the difficulty either: we have an *inner, cognized proposition* that, in itself, possesses all the *semantical* content that is needed, and the fact that it is not publicly accessible does not change this. The idea of replacing propositions with self-attributed properties does, but shows further shortcomings. It seems much more natural to assume that, if clauses such as "... (that) she herself is at home" do express a proposition of a sort – *always intended as a obliqua oratio correspondent to a first-personal proposition* – it is the same as a limited-access proposition (but for the fact that it is not expressed in *oratio recta*), that is a tokening by "I am at home" by, say, Sally, a *fictional* character also affected by amnesia, or "I am loved by all" by John, in a *real* situation. If, in a tale, Sally is described as feeling herself back at home, this is not reducible, as we shall see, to "Sally is at home" and to the "standard" proposition expressed by it.

A last alternative would be the two kinds of propositions approach, but we already discussed that. Notice that, even in the standard view, the that-clause, as in "It is known that Sally is at home" (or, in the real case, that John is loved by all), does not express at all an incomplete proposition, nor two propositions or more. The distinction is not syntactical, it is syntactico-semantical, and finds confirmation in pragmatic usage. What is expressed is a third personal, accessible-to-all proposition, from a similar point of view. So Castañeda's theory (in the rough form of a limited access propositions approach)[7] still seems the more elegant and simple way of understanding such problems.

In Lewis's construal (one is tempted to call it the Quine-Lewis-Chierchia view) subjectivity reduces, roughly, to a point of view, which is void and always, in a sense, one and the same – the general, objective, abstract property we saw above –, with the difference that it corresponds now to a world centered on one individual, now to a world centered on another. In one case (ours) it is centered on John, in another one, say, on me. But there is no real self-knowledge to speak about, only the different descriptions corresponding to different centerings of one world, which is as purely objective as the individuals in it: this is why I say it is the same void point of view that may be attributed, in variants of the objective description, to different individuals, seen *de re*. But in the *Fregean* sense, (*de se*) "thoughts" are PROPOSITIONS (limited-access ones, according at least to some interpreters of Frege, such as Perry).[8]

In Lewis they are all, so to speak, parts of paper worlds, and the objectivity mentioned above is, in my view, lame. Modal realism in this perspective is a way of saying that this world here is *the* world just by chance, not because (real, not fictional) subjectivity *has* a real epistemic grasp of it; and it exists as much, or better as little, as all other "real" worlds. Kripke, I think, has this in the back of his mind when he prosecutes implicitly his old polemic against Lewis in his paper *The First Person*, concluding that first personal access is not just a matter of speaking about objects and persons, but the source of that: «[...]each of us does have a special acquaintanceship with himself or herself, as philosophers from Descartes to Frege have held. This self-acquaintance is more fundamental than anything purely linguistic, and is the basis of our use of first person locutions. And each of us can use them to make genuine claims, to express genuine propositions».[9]

We have reached the heart of the detachment problem. Condition (3) of CDK seems to fail in case we admit sentences with quasi-indexicals ("he, himself" and the like – i.e. Castañeda's quasi-indexical "he*") as *irreducible*. It cannot be valid in its normal form.[10]

If there are such irreducible sentences, they are "ineffable" or "unspeakable",[11] as they are essentially seen from a particular viewpoint, the one of the *oratio recta* of the first person, to which an attitude is thus attributed. It is not attributed from the point of view of the speaker, nor from any objective, third personal point of view. The external attributer does not enter the scene. So such sentences, and even more limited-access propositions that correspond to them, if there are any, are *truths completely internal to a point of view*. We call them "non-detachable truths", and a whole register of language at least, fictional language, is constituted by them (though they are not exhausted by it). A non-detachable truth is a truth that can be known, but does not allow us – starting from the assumption that it *is* known – to derive any sentence (and/or a corresponding propo-

sition) relative to such a truth itself.

That there are such truths, and that they are even expressible, but only so to speak approximately and in *oratio obliqua*, is in itself an argument in favour of an "idealist" or "anti-realist" attitude: a strong form of realism – as so often taken for granted even in the study of propositional attitudes, and doxastic and epistemic stances – would require that every truth be, or be able to be, independent of cognition, hence always re-expressible from a third person point of view – exactly what is not possible here. This is already an interesting conclusion in itself. So far as I know, it is the only *formal* argumentation of this purely linguistic type against materialism that is known: it depends solely on CDK and the factivity of "know".

But a more philosophical example of this is the *cogito* itself. If we choose to read irreducible quasi-indexicality into it, that is. Let us use Hintikka's terminology at first, while stressing quasi-indexical elements.[12] In the performance of the *cogito* the epistemic subject is given to itself, in the sense that it knows that it, itself exists: that I exist. But in this way it is given (I am given) only from a purely "perspectival" point of view for identification. (By acquaintance only, in Russell's terminology as used by Hintikka.) As a consequence, it follows in no way that the epistemic subject – so defined and attainable by itself – exists in any objective, third personal form as described by public means (e.g. the existence of my "real" person as indicated by my name – under knowledge by description).

So far Hintikka, as I understand him is saying: to some the *cogito* is a fallacy, without proof of a substantive metaphysical thesis about the existence of res cogitans, to others it does describe the objective being of the *res cogitans*. In this perspectivalist perspective, both are wrong. We must rather say, now leaving Hintikka aside, that the *cogito* makes known to the epistemic subject a "non-detachable truth", because the sentence itself is an unspeakable one (Banfield), and the singular term in it is a quasi-indexical pronominal that

is possible only in *oratio obliqua*. Idealism in the strong sense reaches only this far.

Castañeda's systematic discovery of quasi-indexicality has its precedents in philosophy (P. Geach, S. Shoemaker), where it developed in formal semantics (from Boër and Lycan to Schlenker), and in reflections inspired by Chomsky on the structures of fiction.[13] Ann Banfield, in this line, showed how a whole function of narration, connected to *discours indirect libre*, is constituted by the use of "unspeakable sentences".

For example, let us consider a tale we can spin in free indirect discourse about Sally, an amnesiac on a journey to retrieve her memories. Fictionality adds an element of indeterminacy to the facts and thoughts we are entitled to glean from the narrative (I use an * to draw attention to explicit or implicit quasi-indexicality):

(4) Sally knows she* is back home

("home", as opposed to "house"). And

(5) Sally is climbing the *stairs toward the* old little room.

Almost all contents are referred back to Sally's point of view. It is – at least *now* and *to her* – *her* old house with *her* stairs; so is the *room*, and the room is old, and possibly little *now*, *to Sally*. So, (i) an entire register of language does not have the aim of communicating objective information, but of expressing subjectivity for the subject itself; (ii) this register is, to some extent, present everywhere, not just in literature but also in general narration or explaining. In literature, it may happen that some retrieved, and so rightly attributed, memories, even turn out to be false.

Non-detachable truths with such a vast and various scope do perhaps pose a limit to testimony too, as well as to first-person authority and their/its ability to transmit knowledge. They also seem to pose limits on the centrality of communication.

What the case of non-detachable truths generally shows is that the most "internal" mode of knowledge (that about oneself) poses a highly interesting threat to the logical expressibility of knowledge, and CDK. (Some would be tempted to say that *de se* true beliefs about oneself fail to be knowledge for logical reasons: it is not, to say the least, a logically real and proper case of propositional knowledge.) From one point of view, these facts determine a limit to externalist approaches. Self-knowledge is an internal phenomenon, at least in large part. On the other hand, the threat *is* similar to that posed by externalist theories themselves, that place knowledge conceptually before justification. But we can ask: why should we have expected otherwise? If we accept that self-knowledge is knowledge metaphysically, why should we suppose that knowledge always comes in the form of a given true p that is known by an epistemic subject, and as such can be known by another one too?

In self-knowledge, we have knowledge as a unity, not as a given p that can be known by this or that, or a third, subject. If we accept this, it is no longer so surprising that, *as logically defined knowledge*, knowledge about oneself is so difficult, or maybe impossible, to pin down in the usual ways, both for the internalist and the externalist, both as a result of analysis, and as a result of successful, but always partial, transmission through testimony (it exists, but it is to be taken for granted, in a sense). This general difficulty is perhaps not so great in view of the fact that self-knowledge is, in itself (though it could be in reconstruction), in no way "objectual" knowledge.

Back to the Sally example. We have here, in fiction, a special case. In "Sally knows she is home" *we* cannot detach, roughly said, "she is home". (Also because not all the alleged facts must be real, in fiction.) But *we* can detach, it seems, "Sally is home", can we not? (*In case we actually had the complete construction with the subordinate clause at our disposal*). *Ceteris paribus*, and in the general case, however, it can even be doubted that Sally

herself can do this, that even she can go this far. What really is implied is "I am home" as thought by Sally (who might even not be aware that she is Sally). Does this mean that she *really* is home?

From her point of view, in principle, it is not to be assumed that Sally herself is able to detach "Sally is home". But the point goes deeper. Is the fact that Sally knows that she, herself is home the same fact as knowing that Sally is home? More precisely: that she herself is home in her first personal point of view is an experience, so it can be the same as a fact (that Sally, or a person publicly identified under such a name, is home) only from a perspective. It cannot express the same proposition as "Sally is at home".

If "... she, herself is at home" should express – at least to the outside observer – the same proposition as "Sally is at home", then the latter would no doubt be detachable. But this would condemn all the convincing analyses we made reference to, and which make very good sense in explaining immunity from error due to misidentification (Shoemaker), types of anaphora, quasi-indexicals (Castañeda), uses of *discours indirect libre* in fiction (Banfield, Schlenker), and, today, even Kaplanian monsters (Higginbotham and Schlenker, though in different ways).[14] Why this would wipe out this whole sector of research seems clear, but it bears repeating. It also has autonomous argumentative value. An *exclusively* embedded sentence, a subordinate clause, does not seem able to express the same proposition as a main clause.

Many scholars tend to treat the problem of quasi-indexicals in a partial way, because they stick to tradition and reject limited-access propositions,[15] or because they make a misleading distinction between semantic and cognitive propositions, applying it across the board.

We can rather suppose that all sentences of the type "I see blue" express propositions containing elements with a qualitative character, or have *qualia* in them. Propositions expressed by, say, "A sees blue" would be different in nature. Just given the similarity in

the innate constitution of speakers, and the similar way in which they learned language (i.e. also by ostension), I and almost everyone else will always agree on by far most tokenings of, say, "this is blue" expressed by me or them.

As a matter of fact, the only constitutively cognitive propositions would be the ones with "I", that is exactly those that are expressed by quasi-indexicals in *oratio obliqua*. In "Sally knows that she herself is home", "she herself" stands for the occurrence of "I" in such a proposition. So, as widely agreed in the literature, though not always clearly spelled out, the embedded clause cannot be synonymous with "Sally is home", which is detachable.

▌ Conclusion: The cogito and the non-detachment problem

First a response to a possible criticism. Someone may object, however, that from the fact that Sally knows that she herself is home, it *does* follow that Sally is home, as it follows from Sally veridically tokening "I am home". Now, first of all, in fictional discourse at the very least this is far from obvious: that could be a case of *déjà vu*, or illusion, much as in the above examples it could be a matter of personal *qualia* that are different from objective, and intersubjective, properties of reality (so it may be *true* that I *see* purple where you *see* a shade of dark pink) – and yet Sally could experience a true feeling to herself, the feeling of being at home. This is, however, a moot question.

For secondly, and more importantly, if it follows that Sally is at home, this does not just follow from the fact that the non detachable phrase "... that she herself is at home" (*as if it were meaningful and possible* to write it down this way, *qua* detached), is a known truth, *nor from the fact* that "Sally knows that she herself is at home" is a known truth. It follows from this *plus* the objective, intersubjective and third personal identification of the person who could, in this case, say "I"

with the *objective Sally*. We come full circle, to what was exactly our point: it does not follow directly in the way in which the truth of 2+2=4 follows from the known fact that John knows that 2+2=4 (a knowledge in which there never can be any talk about *qualia* and the like).

Thirdly, we mean that, detachment being impossible, the derivation is impossible *per se* and without further assumptions, not that these further assumptions are impossible and not obvious, as they actually are in many cases (uninteresting ones, usually). The objection is commonsensical, but begs the question.

Lastly, in the case of fiction, or in general when it is said, but not (publicly) known, that A knows something of himself or herself, reaching objectivity from a third-personal viewpoint seems even more difficult. This has to do with the fact that the fact that A knows p does not always entail that it is known that A knows p: it may just be a reasonable presupposition, or a report about an utterance (say, about toothache). The question-begging in the usual objection(s) can be even more widely misleading than one suspects at first sight.

All this is seen in the *cogito* too (in the form "I think, therefore I am", or better "this sentence 'I am, I exist [as a thinking thing or substance]' is necessarily true every time I pronounce it or conceive it in the mind"). In Hintikka's first interpretation, there is no formal derivation, just a performance with presuppositions. Many people, including Hintikka, would say that this is not enough to prove that I exist as a thinking substance, or even a thinking thing. There is also, as previously mentioned, a tradition, in analytic philosophy, of scholars denying the *cogito* outright (Carnap, Ayer). This is why I said that to some the *cogito* is, in short, a "fallacy". To others, it is, traditionally, the proof of the autonomy of thinking. But Hintikka's final interpretation sets things aright.[16]

If I understand him, he claims that what is referred to in the *cogito* as subject is not referred to in a public way, but from "perspec-

tival", inner knowledge (a bit like Russell's sense-data). So, in our construal, the *cogito* unpacks as "I think, therefore (I know that) I, myself exist", or more simply "I know that I, myself exist". This is also (in the second, embedded clause) non-detachable knowledge: I can prove that I, as perspectively seen by myself, exist, but *not* that the present writer, as publicly recognized, exists, even less as a thinking substance among other objective, intersubjectively given substances. If I say something, by the performance itself I exist, but, again, not in any specific, given way (yet). So the *cogito* is not a fallacy, but it is not the proof of the objective existence of something *qua* something *either*. This requires public identification and objective research.

Let us sum up our results. These are cases in which, from given (subjective) knowledge, one does not reach truth; so that truth cannot always be offered, by testimony or other means, to be developed into knowledge by further subjects: communication proper is, in part, impossible.

To say the least, this is a final proof that there are forms of testimonies (or even oblique communication, if you so prefer) that may be taken as truths, but basically do not have any role in the search for objective truth, and intersubjective knowledge. If we are right, non-detachability means that the truth condition cannot be applied as a criterion here,[17] so that even in this respect there is an insufficiency of CDK. As we have seen, not only does this approach throw some new light on the validity of the *cogito*, but it can help us construe (section) a purely linguistic argument against radical realism, and the naturalistic idea that all contents can be described from a third-person point of view.

It is finally of great interest that the (this) most "internal" mode of knowledge about oneself (which makes up the *cogito* too, with its conclusion, "ergo sum": a non-detachable form of "inner" *de se* knowledge) poses a threat to the logical explicability and expressibility of knowledge, and to CDK. This threat is – though internalist, and such as to

suggest that we should spare with the truth condition –, also akin to that of externalist theories, which place knowledge conceptually before justification.

It also suggests that knowledge about oneself is not just a matter of linguistic description of pre-given knowledge, but of a wholly different dominion of knowledge or at least consciousness, prior to language, to be brought to the attention of the philosopher of mind, and language. It is surprising and ironic that knowledge about oneself, so direct, is (in its necessary expression) so difficult logically to pin down with respect to other dominions of knowledge. This is only partly explained – partly following Galen Strawson, especially as far as diachronic aspects are concerned[18] – by the seeming fact that it is in no way properly an "objectual" knowledge of a self or selves, but rather a particular mode of access to facts both external and private.

Acknowledgements

Notes

[1] For a definition of *de se (self-referential* attitudes) see D.K. LEWIS, *Attitudes de dicto and de se* (1979), reprinted in: D.K. LEWIS, *Philosophical Paper*s, vol. I, Oxford University Press, Oxford 1983, pp. 133-159, with a "Postscript 1983".

[2] From a strictly linguistic point of view, there are no problems in translating this sentence ("John knows that John is loved by all") in any language. To translate "John knows that he, himself is loved by all" is more complicated. In this case, languages such as Italian or Spanish use infinitives, whereas English, for instance, remains ambiguous, so that the corresponding English sentences are not fully adequate. I make this remark because Gennaro Chierchia argued convincingly that (in sentences such as "Giovanni sa di essere amato da tutti") the *de se* reading is compulsory, see G. CHIERCHIA, Anaphora and *de se* Expressions, in: R.R. BARTSCH, J.F.A.K. VAN BENTHEM

(eds.), *Semantics and Contextual Expressions*, Foris Publications, Indianapolis 1989, pp. 1-31. This means that, on a first approximation, such sentences exclude replacement by any subject other than the tacitly understood subject, and express something that could be called direct self-consciousness. Sentences such as: "Giovanni sa che egli è amato da tutti" by contrast do not have a compulsorily *de se* reading it seems, even in the coreferential case - just as in English. Infinitives are possible in English with verbs such as "to expect". But I have no intuitions about the *de se* reading in such and similar cases.

[3] See, i.a., F. ORILIA, W.J. RAPAPORT (eds.), *Thought, Language and Ontology. Essays in Memory of Hector-Neri Castañeda*, Dordrecht, Boston, London, 1998; W.J. RAPAPORT, S.C. SHAPIRO, J.M. WIEBE, *Quasi-indexicals and Knowledge Reports*, in: «Cognitive Science», vol. XXI, n. 1, 1997, pp. 63-107.

[4] See R. MORAN, *Authority and Estrangement: An Essay on Self-knowledge*, Princeton, Princeton University Press 2001.

[5] M. KAMERMANS, *An Introduction to Japanese. Grammar Syntax and Language*, on-line version April, 26th, 2010, p. 115.

[6] They are connected to logophoricity too: see G. VARNIER, *Quasi-Indexicals, Kaplanian Monsters, and Self-Consciousness*, in: A. PALMA (ed.), *(Hector-Neri) Castañeda and his Guises*, de Gruyter, Berlin/Boston 2014, pp. 161-185.

[7] For an approach in this sense see G. VARNIER, *Quasi-Indexicals, Kaplanian Monsters, and Self-Consciousness*", cit.

[8] See, also for a discussion of Castañeda's views, J. PERRY, *The Problem of the Essential Indexical*, in: «Nous», vol. XIII, 1979, pp. 3-21. On Frege see R. MAY, *Frege on Indexicals*, in: «Philosophical Review», vol. CXV, n. 4, 2006, pp. 487-516; W. CARL, *The First Person Point of View*, de Gruyter, Berlin/New York 2014. Perry and others reject this solution, however.

[9] S. KRIPKE, *Philosophical Troubles*, Oxford University Press, Oxford/New York 2011, p. 319.

[10] In a way, this has to be taken into account in designing expert systems, and so has some relevance for artificial intelligence, see W.J. RAPAPORT, *Prolegomena to a Study of Hector-Neri Castaneda's Influence on Artificial Intelligence: A Survey and Personal Reflections*, in: F. ORILIA, W.J. RAPAPORT (eds.), *Thought, Language and Ontology. Essays in Memory of Hector-Neri Castañeda*, cit., pp. 345-367.

[11] See A. BANFIELD, *Unspeakable Sentences. Narra-*

tion and Representation in the Language of Fiction, Routledge & Kegan Paul, Boston/London/Melbourne 1982.

[12] See J. HINTIKKA, *Cogito, Ergo Quis Est?*, in: «Revue Internationale de Philosophie», vol. L, n. 195, 1996, pp. 5-21; J. HINTIKKA, *Perspectival Identification, Demonstratives and "Small Worlds"*, in: «Synthese», vol. CXIV, n. 2, 1998, pp. 203-232.

[13] See P. GEACH, *On Beliefs about Oneself*, in: «Analysis», vol. XVIII, n. 1, 1957, pp. 23-24; S. SHOEMAKER, *Self-Reference and Self-Awareness*, in: «The Journal of Philosophy», vol. LXV, n. 19, 1968, pp. 555-567; S. SHOEMAKER, *Introspection and the Self*, in: «Midwest Studies in Philosophy», vol. X, 1986, pp. 101-120; S. LYCAN, S. BOËR, *Knowing Who*, MIT Press, Cambridge (MA) 1986; P. SCHLENKER, *Propositional Attitudes and Indexicality: A Cross-categorial Approach*, MIT Doctoral Dissertation 1999; P. SCHLENKER, *Indexicality, Logophoricity, and Plural Pronouns*, in: J. LECARME (ed.), *Research in Afroasiatic Grammar II, Selected Papers from the Fifth Conference on Afroasiatic Languages*, Benjamins, Amsterdam/Paris 2000, pp. 409-428; P. SCHLENKER, *Context of Thought and Context of Utterance (A Note on Free Indirect Discourse and the Historical Present)*, in: «Mind and Language», vol. XIX, n. 3, 2004, pp. 279-304; P. SCHLENKER, *Indexicality and De Se Reports*, in: K. VON HEUSINGER, C. MAIENBORN, P. PORTNER (eds.), *Semantics: An International Handbook of Natural Language Meaning*, vol. II, de Gruyter, Berlin/New York 2011, pp. 1-41.

[14] P. SCHLENKER, *A Plea for Monsters*, in: «Linguistics and Philosophy», vol. XXVI, n. 1, 2003, pp. 29-120; J. HIGGINBOTHAM, *On Words and Thoughts about Oneself*, in: F. RECANATI, I. STOJANOVIC, N. VILLANUEVA (eds.), *Context-Dependence, Perspective and Relativity*, de Gruyter, Berlin/New York 2010, pp. 253-282.

[15] See J. PERRY, *The Problem of the Essential Indexical*, cit.; E. CORAZZA, *Reflecting the Mind: Indexicality and Quasi-Indexicality*, Oxford University Press, Oxford 2004; E. CORAZZA, *Essential Indexicals and Quasi-Indicators*, in: «Journal of Semantics», vol. XXI, n. 4, 2004, pp. 341-374.

[16] See J. HINTIKKA, *Cogito, Ergo Quis Est?*, cit.

[17] Notice that only if testimony is taken for granted, and an outside observer is assumed, can CDK be applied as a criterion. If "A knows p", or better we want to be sure of that, we can easily establish that he or she believes and is justified in believing that p, but we need an outside observer to warrant that p is true, so that all the three conditions are fulfilled. What we may call the "Cartesian situation" is exactly a situation where no outside observer is conceivable, and so, among other things, skepticism about knowledge also in the first person seems to follow.

[18] See G. STRAWSON, *Against Narrativity*, in: «Ratio», vol. XVII, n. 1, 2004, pp. 428-452. For his defense of selves see G. STRAWSON (ed.), *Consciousness and Its Place in Nature*, Imprint Academic, Charlottesville (VA) 2006; G. STRAWSON, *Realistic Monism: Why Physicalism Entails Panpsychism*, in: «Journal of Consciousness Studies», vol. XIII, n. 10-11, 2006, pp. 3-31, followed by comments; G. STRAWSON, *The Minimal Subject*, in: S. GALLAGHER (ed.), *Oxford Handbook of the Self*, Oxford University Press, Oxford 2011, pp. 253-278; G. STRAWSON, *The Evident Connexion: Hume on Personal Identity*, Oxford University Press, Oxford 2011.