

Anaphora and Attitudes *De Se**

Gennaro Chierchia

The problem of so called 'essential indexicals' (or, in Lewis's words, of attitudes 'de se') has been widely discussed in the philosophical literature since it was first pointed out in Castañeda's work. Various evidence seems to suggest that the perspective one has on oneself, and consequently the attitudes towards oneself are, in some way, special. The question is to spell out exactly how special. Many proposals have been made in this connection.¹

The question of the nature of attitudes towards oneself, besides being central for cognition in general, has an obvious specifically semantic relevance. From a semantic point of view, the issue is the structure of the logical form (or truth-conditional import) of reports of attitudes towards oneself. For example, one might ask the following questions: can the semantics of such reports be reduced to the semantics of ordinary (i.e. non de se) attitude reports? Or are there special ways in which relations towards self are grammaticized?

These are the issues that I would like to address here, in a tentative way and without pretense of exhaustiveness. The wide variety of proposals on essential indexicality have precipitated very little in terms of semantic analyses of attitude reports, perhaps not surprisingly, given the complexity of the topic. With few exceptions (such as e.g. Cresswell (1985)), it is not obvious how most of the currently available approaches to essential indexicals should be related to an interpretive procedure for belief sentences. In fact, many authors have been very explicit in warning against connecting directly the metaphysical issue of the object of mental attitudes to the problem of the semantics of belief sentences (see e.g. Lewis (1979, p. 541), Stalnaker (1981, fn 16)).

My general goal here will be to outline a semantics for de se attitude reports based on a version of Lewis's proposal that properties (rather than propositions) are the objects of attitudes towards oneself. I will try to argue that this view has interesting linguistic consequences. I will also offer some reasons for being skeptical about the explanatory power on

* I wish to thank A. Bonomi, B. Partee, and T. Reinhart for their comments. Special thanks are due to J. Higginbotham for his thought-provoking comments as I presented this material at MIT in October 1987 and to R. Stalnaker for stimulating and very helpful conversations on it. For any problem and confusion, I alone am responsible.

this score of 'propositional' theories of *de se* phenomena. This will show, I hope, that philosophical and linguistic considerations can fruitfully interact in shedding light on how language works.

The organization of the present paper is as follows. First (in section 1) I will outline what I take the problem to be and informally sketch the solution that I have in mind. I will subsequently present (in section 2) more in detail the grammar that I envisage. Finally (in section 3) we will consider some empirical consequences of such grammar.

1. REFERRING TO SELF

1.1. The problem

Consider the sentences in (1)

- (1)
- a. Pavarotti believes that the one who can sing 'O che gelide manine' without mistakes is a musical genius
 - b. Domingo is the only singer that can sing that aria without making mistakes
 - c. Pavarotti believes that Domingo is a musical genius

It is standard to interpret (1a) in two ways. On the one hand, Pavarotti's belief might be, as it were, purely conceptual. He is not directly acquainted with the singer in question. He believes that whoever that singer turns out to be is a genius. On the other hand, there might be some kind of link that directly relates Pavarotti to the person that in fact is the singer in question. For example, Pavarotti might be perceptually familiar with (and therefore causally linked to) him. The first type of belief is usually called *de dicto*, the second *de re*. Generally, the test for *de re* belief is taken to be substitutivity of coreferential terms. Grossly oversimplifying, if Pavarotti's belief is *de re* and (1b) is true, one could also report Pavarotti's belief as in (1c).²

The characteristic of a *de re* belief is that it involves a 'strong' epistemic relation of belief-bearers to the objects of their beliefs. Now, the epistemic access we have to ourselves is certainly a 'strong' epistemic relation. The question that arises, then, is whether the latter form of access has a somewhat special status, that singles it out from the other *de re* beliefs.

There are various facts that have been argued to show the special character of attitudes towards oneself. The following, a variant of an example due to D. Kaplan, is illustrative of the kind of relevant phenomena. Consider the sentence in (2).

- (2) Pavarotti believes that his pants are on fire

Sentence (2) can be true in two distinct sets of circumstances. It can be true in a set of circumstances where Pavarotti is aware that his belief concerns himself and in a set of circumstances where he is not.

To see the latter, imagine Pavarotti looking at a mirror without realizing it and seeing a man whose pants are on fire. The man Pavarotti is seeing in the mirror is in fact Pavarotti himself, but he does not realize that.

What is puzzling about this situation is the following. There exists a specific entity whom Pavarotti has a belief about, namely himself. He is in fact seeing someone and his beliefs are causally determined by the man he is seeing; his dispositions would be systematically affected, if that man's properties were different from what they are. Hence, Pavarotti's belief is a *de re* one about the man he sees. Yet, Pavarotti is not aware of the fact that his own pants are on fire. His mental state is not the one of someone who says to himself 'my pants are on fire' circumstance which affects his dispositions in significant ways. Only the awareness that the man he is seeing is himself would prompt him to run for the fire extinguisher. This does seem to at least suggest that the access we have to ourselves is qualitatively different from the access (however direct) we have towards a certain 'res'.

From a semantic point of view, the above facts might be taken as evidence that sentences such as (2) are ambiguous.³ Someone who utters (2) might either be reporting that Pavarotti has a belief about a certain *res* (that, unbeknownst to Pavarotti happens to be Pavarotti himself); or else s/he might be reporting a belief that Pavarotti has in full awareness of his role as the object of such belief. On the first reading, sentence (2) would be verified by the pants-on-fire-plus-mirror situation just described. Furthermore, on such a reading, if Pavarotti were questioned about his beliefs, he would typically not report them using the first person pronoun 'I' or 'my'. On the second reading, sentence (2) would be false in the pants-on-fire-plus-mirror situation. For the second reading to be true, if Pavarotti is questioned about his beliefs, he would have to report them using the first person pronoun. This second (hypothetical) reading of (2) is the one that we call, following Lewis, '*de se*'. Throughout this section I will use *de se* and *de re* as descriptive labels for the particular kind of contrast just discussed.

So, one way of analyzing our Kaplanesque example is by taking (2) to have two readings. This ambiguity could be resolved at some level in the grammar of English by assigning to (2) two distinct logical forms. Such logical forms while non equivalent, will of course not be independent of one another. For the kind of epistemic access that characterizes attitudes towards oneself will certainly sustain attitudes towards a certain *res* (namely,

the self). So the two hypothetical logical forms used in expressing the de se and de re sense of an attitude must be chosen in such a way that the former entails the latter.⁴

There is, however, another very plausible way of analyzing the example in question (defended in, e.g., Boer and Lycan (1980)), which goes roughly as follows. De re beliefs are notoriously problematic because, among other things, we can access a certain entity from different perspectives that are equally 'direct' or 'vivid' and therefore equally suited to license beliefs about specific res. But different ways in which an object is presented can lead one to attribute to it contradictory properties, without thereby necessarily condemning one to having contradictory beliefs. A famous case in point is Quine's Ralph who sees Orcutt in two guises. Under one guise, Ralph believes of Orcutt that he is a spy, under another guise that he is not. Our Kaplanesque example discussed in connection with (2) might be taken to be an instance of this problem. In the situation of the example, Pavarotti has access to himself in two different ways: as the person that he would refer to using the pronoun 'I' and as the person he sees in the mirror. Both ways of accessing himself may arguably make his attitude towards himself de re (and/or license an illocutionary agent to report his attitude as de re). Yet under one perspective he believes of Pavarotti that his pants are on fire, under the other he doesn't. Thus the problem of (2) reduces to a more general problem about de re. Any solution to the latter, one might contend, will extend to the former. No "special" logical form is needed beyond what we need to cope with standard de re phenomena.

In spite of the prima facie strong plausibility of this second line of approach, in what follows I will pursue the first one and try to make a case that:

- (a) de se readings are associated with specific linguistic structures
- (b) the compositional semantics of such structures crucially resorts to properties rather than propositions.

The idea that attitudes de se must be understood as relations of agents to properties originates with Lewis (1979), and I shall start by briefly reviewing his proposal.

1.2. A solution

Lewis, by means of a series of thought-experiments (e.g. the famous example of the identity crisis of two gods), argues that one could be propositionally omniscient, while still lacking crucial information about self. For example, for any object α and any property P we might know whether α has P, without knowing, say, which of those α 's we are. We could know all the

true propositions and still be unable to locate ourselves among the inhabitants of a certain world.

Lewis concludes from this that propositions lack the structure needed to classify essentially indexical mental attitudes. As a way out, Lewis develops what one might call a 'self-locational' view of belief. He proposes that to believe that ψ is to believe oneself to have the property of inhabiting that region of the logical space where ψ holds (to be in a world where ψ). Different attitudes can be viewed as various ways of locating oneself in logical space. But, Lewis argues, there is no reason to maintain that mental attitudes are limited to locating oneself in *logical* space. We better be able to locate ourself also in physical space-time. The latter kind of self-locating belief is what is at the basis of essential indexicality. For example, if I attribute to myself the property of being Italian, I thereby situate myself in a region of our space-time where that property is satisfied by me. Roughly put, self-attribution of a property that doesn't differentiate among the inhabitants of a world corresponds to a de dicto belief (locating oneself in logical space), self-attribution of a property that does to a de se belief (locating oneself in actual space-time). This is, in a very condensed form, Lewis's proposal.

Now, Lewis's proposal rests on thought experiments that tend to show that propositions do not have enough structure to represent essential indexicality. That is of course going to depend on what notion of proposition one is working with. In particular, Lewis takes propositions to be sets of worlds. Perhaps a more richly articulated view of propositions might give us good ways of representing attitudes de se. Indeed, this has been argued for by many. But even if one could, in principle, develop a notion of proposition that enables us to characterize abstractly attitudes de se, it might still be the case that natural language uses properties for this purpose. And this is precisely what I think happens.

To see why, we must first bite the bullet and translate Lewis's proposal into a claim on the meaning of attitude reports. As a first shot, we might assign to sentences such as (2) above the representations given in (3).

- (3) a. believe (P, λy [y is in a world where P's pants are on fire])
- b. believe (P, λx [x is in a place where x's pants are on fire])

The believe-relation in (3) is a self-ascriptive relation, roughly paraphrasable as 'believe oneself to have'. Self-ascribing the property in (3a) amounts to believing that the actual world is located in a region where Pavarotti's pants are on fire. The property in (3a) does not discriminate among the inhabitants of a world: it is a constant property. The property in (3b), instead, does. In order to self-ascribe the property of being an x such that x's pants are on fire, we have to be located (or believe ourself to

be located) in some specific spatiotemporal region where that property is satisfied by us.

Furthermore, if Pavarotti is related by a 'vivid' or 'strong' acquaintance relation to the object of his belief, his belief would be *de re*. The *de re* counterpart of the belief in (3) might be represented as in (4).

- (4) $\lambda x[\text{believe}(x, \lambda y [y \text{ is in a world where } x\text{'s pants are on fire}])(P)$

If proper names are rigid designators, then (3a) and (4) become equivalent. In the following discussion, however, I do not assume that they are. As far as I can see, the central aspects of the proposal to be developed do not hinge on this assumption or, for that matter, on any specific assumption concerning the logical form of attitudes *de re*. Readers should feel free to substitute for (4), their favourite representation of *de re* beliefs.

Be it as it may, Lewis's view construed as a semantic claim might appear rather bizarre, as it leads one to interpret that-clauses as properties. The strangeness of this claim rests, I take it, on the fact that it somehow would clash with our pretheoretic intuitions.

One could, of course, point at the fact that intuitions about the meaning of that-clauses are always theory-laden. But at any rate, we don't have to be so radical as Lewis. I think that an intermediate position between the standard one and Lewis's could and should be maintained. We might, for example, assume that in general that-clauses denote propositions, but in certain cases they denote properties, namely when *de se* reports are involved. This would mean that, putting aside *de dicto* constructions, a sentence such as (2) would have the two logical forms shown in (5).

- (5) a. $\lambda x[\text{believe}(x, x\text{'s pants are on fire})(P)$
 b. $\text{believe}(P, \lambda x[x\text{'s pants are on fire}])(P)$

In (5a), *believe* is a relation between Pavarotti and a certain proposition, one that turns out to be about Pavarotti himself. In (5b) *believe* is a self-ascriptive relation between Pavarotti and a property. We assume that relations of mixed types (i.e. proposition- and property-taking) do exist. We could justify this 'mixed' theory of the nature of the belief-relation in various ways. One might be shifting from Lewis's 'self-locational' picture to what Stalnaker (1984) calls the 'pragmatic' picture of mental states. In very rough terms, according to the latter when agents engage in actions, they are typically confronted with a range of possible outcomes and they have attitudes, *pro* and *con*, such possible outcomes that affect their choices and dispositions. Propositions are ways of classifying the alternatives that we are confronted with in action and thus it is very reasonable to try to classify mental attitudes as relations to propositions. Now, properties

can be thought of as recurring patterns that obtain across different ways in which the world can be. For example, the property of being 6 ft tall classifies the state that, in a world *w*, some *x* is in just in case *x* is 6 ft tall in *w*; the property of running classifies the type of action that *x* undertakes in a world *w* just in case *x* runs in *w*. We have attitudes *pro* and *con* possible outcomes of our actions, and likewise we may have attitudes *pro* and *con* types of states and types of activities that cut across such possible outcomes.

Now, as pointed out above, by its very nature a relation towards self links us directly to a certain *res*. Consequently, the truth of a *de se* belief-report will entail the truth of a *de re* belief-report. For example, (6a) entails (6b) but not viceversa:

- (6) a. Pavarotti believes that he, himself has pants on fire
 b. Pavarotti believes that that person (pointing at Pavarotti) has pants on fire

The emphatic reflexive in (6a) is meant to force us to select the *de se* reading of (6a). Furthermore, I assume that demonstratives and personal pronouns are generally associated with a *de re* interpretation. Given the line I am taking on how *de se* vs *de re* attitudes should be represented, it is reasonable to take this pattern of entailments as criterial for 'self-ascriptivity'. I.e. a relation of *x* to a property *P* is self-ascriptive iff it entails a (*de re*) relation to the proposition $P(x)$, without being entailed by it. This is the definition that we will be assuming in what follows.

However, I regard the above definition of 'self-ascriptivity' as highly provisional. Depending on what one assumes on attitudes *de re*, there are various plausible alternatives to it. In particular, there are two main families of strategies that are directly relevant in this connection and perhaps it is appropriate to discuss them briefly.

The first is to take, following Lewis, self-ascription as a primitive concept and use it to characterize attitudes *de se*. Attitudes *de re* can then be defined as a special case of attitudes *de se*. In developing a theory of logical form based on this view, the desired entailment should fall out of such a definition.⁵

The second strategy is based instead on the development of a theory of attitudes *de re* and the attempt to define attitudes *de se* as a special case of the latter. To give some substance to this proposal, let me sketch two ways of developing it.

One possibility might be to adopt a definition like the following:

- (7) *x* stands in the belief relation with property *Q* (i.e. *x* self-ascribes *Q*, in Lewis's terms) iff *x* believes (*de re*) that *x* has *Q* and furthermore

$K(x,x)$, where K is the cognitive access that we have to ourselves.

The definition in (7) (which is meant only as very rough suggestion) presupposes that we can make independent sense of (a) *de re* belief and (b) of the relevant cognitive relation.

As far as (a) goes, I have nothing to say. As far as (b) goes what one might want is something like the following:

$K(x,x) =_{df}$ x is disposed to describe the relevant belief by referring to x by means of the first person pronoun.

This however would make the definition in (7) dependent on a particular property of the representation system of the believer, which might render problematic the ascription of *de se* beliefs to creatures with limited symbolic capacity (and yet arguably capable of *de se* attitudes), like birds or sharks.

In fact, the proposal in (7) is not far from one developed by Max Cresswell. Cresswell (1985, ch. 14) builds a semantics for discourse *de se* that uses speaker-centered worlds (developing an earlier proposal by Quine). Roughly, his semantics is such that, for example, Pavarotti believes *de se* that his pants are on fire iff in all his belief-alternatives centered on him (i.e. for all $\langle w,P \rangle \in H_p$, where H_p characterizes Pavarotti's belief-alternatives) his pants are on fire. A Pavarotti-centered world is to be understood, intuitively, as a world where Pavarotti is the speaker.

Formally Cresswell's proposal might do the job. That is, it provides us with a richer notion of proposition that has enough structure to represent attitudes *de se*. In fact, it turns out that speaker-centered worlds are in a sense isomorphic to properties, so that there is a way of translating back and forth between Lewis's proposal and Cresswell's.⁶

However, such a proposal faces objections similar to those faced by (7). For even if we understand the notion of speaker very broadly, it is hard to see how it could be applied to elementary creatures that might, nevertheless, have *de se* beliefs. What one would really like is a notion of "self"-centered world. But this notion *per se* is not much clearer than the notion of self-attribution or, for that matter, of the notion of cognitive access to self, used in (7).

An alternative story, in a similar vein, might go as follows. As Quine, Kaplan and others have argued a *de re* belief is always relative to a particular perspective. This raises the issue whether there exists something like an "absolute" perspective on things. It might be suggestive to pursue the idea that the only thing that we can access from such an "absolute" point of view is ourselves. Only to ourselves we are linked in a purely *de re* (i.e. non perspective-dependent) mode. If this turns out to be viable, one

could then adopt a definition slightly simpler than (7), along the following lines:

(8) x stands in a self-ascriptive relation R to property Q iff $R(x,Q(x))$

The definition in (8) presupposes that we independently understand attitudes *de re* (in order to make sense of the right hand side), but does not make appeal to "cognitive access to self" as a primitive.

Any of the approaches just sketched could easily be implemented in such a way as to yield as a consequence that (5b) entails (5a) and could be used to provide a semantics that would account for the pattern of entailment illustrated in (6).

I don't know which of these strategies is the right one. The issues that are involved in choosing among them are some of the hardest issues surrounding our understanding of propositional attitudes. I think that my proposal is independent of finding *the* solution to these issues (to the extent that some solution is forthcoming). What I regard as crucial for my proposal is that attitudes *de se* should be construed as relations of agents to properties that involve, indirectly, a predication or ascription of that property to the agent in question, where this indirect form of predication manifests itself in patterns of entailments such as (6) above. I shall argue that this is the way that natural language works. But before turning to a discussion of the relevant evidence, I must show how this idea about attitudes *de se* can be worked into specific semantic theories, i.e. how the logical forms in (5) can be compositionally assigned to the relevant English sentences.

2. THE GRAMMAR OF *DE SE*

In general, the approach sketched above requires one to grant that the relation between syntactically clausal structures and semantically propositional entities is more complex than what one might be led *prima facie* to believe, a point made, on independent grounds, in Higginbotham (1986).

The basic idea is that sentences containing a pronominal element can act as "open" formulae or unsaturated structures (i.e. properties). This links the possibility of a *de se* reading of a clause specifically to the presence of a pronominal element. It is customary to assume that pronouns like *she* or *he* can be interpreted as variables that either get their values from the context or are bound by some quantifier (in the right structural configuration). Now, we can imagine that pronouns can also optionally be bound by, say, a property abstractor. I will sketch two ways of compositionally implementing this idea, one in a Government and Binding-

like framework, the other in an extended categorial grammar. (Some familiarity with the basic concepts of these two approaches will have to be presupposed).

2.1. A GB-approach

If we assume that S(urface)-structures are mapped by rules of construal into L(ogical) F(orms) for which a truth-conditional interpretation is specified, we might say that pronouns can either be bound directly by a C-commanding NP or indirectly via an operator adjoined to S' (S-bar or C(OMP) P(hrase)). So, for example, (9a) can be associated with the two LF's given in (9b) and (9c):

- (9) a. Mary thinks that she is in danger
 b. $Mary_i$ thinks that [she_i is in danger]
 c. $Mary_i$ thinks [O_i that [she_i is in danger]]
 e. think (M, $\lambda x[x$ is in danger])
 d. $\lambda x[\text{think}(x, x \text{ is in danger})](M)$

Operators must be locally assigned a range, i.e. have an antecedent (or be controlled) whence the coindexing between *Mary* and O_i in (9c). Such a coindexing yields the desired effects in terms of the behavior of the pronoun with respect to opacity phenomena (and with respect to agreement between *she* and its antecedent). The operator in (9c) should be understood as a λ -abstractor which means that (9c) is interpreted as in (9e). This gives us the de se reading of (9a) and it contrasts with the LF given in (9b) which is interpreted as in (9d).

The very same contrast obtains, of course, if we have quantified NP's. Thus consider, for example, (10a).

- (10) a. Everyone in that room thinks that he is Hume
 b. [[everyone in that room]_i]_i thinks that [O_i [he_i is Hume]]]
 c. $\forall x_i[\text{person in that room}(x_i) \rightarrow \text{think}(x_i, \lambda x [x \text{ is Hume}])]$

The most plausible interpretation of (10a) claims that each person in the relevant room has a certain de se attitude (perhaps due to schizophrenia). This interpretation is associated with the LF in (10b), associated with the truth-conditions represented in (10c). The fact that the de se/de re contrast shows up in quantified statements shows that it cannot be reduced to the standard kind of bound anaphora. One kind of binding cannot yield two different interpretations. A different binding device, such as an abstractor, is called for.

It should be noted that the relation between a pronoun construed de

se and its antecedent, is not subject to so called island constraints, as shown in (11).

- (11) a. $John_i$ thinks O_i [that Mary is still wondering [whether to marry him_i]]
 (11) b. $Bill_i$ believes O_i [that the fact that people like him_i is a miracle]

Clearly, the pronouns in (11) can be understood de se. Thus, we should presumably assume that the operator in question is base generated rather than adjoined to S' by movement.

It is perhaps worth pointing out, in this connection, a proposal brought forth by Williams (1977) to deal with sloppy readings in VP-anaphora phenomena, such as the one illustrated in (12a).

- (12) a. John likes his brother and Mary does too
 b. [VP likes his brother] $\Rightarrow O_i$ [VP likes his_i brother]
 c. $\lambda x[x$ likes x's brother]

The reading of (12a) relevant here is the one according to which Mary likes her own brother. To get such reading, Williams proposed a "pronoun rule" whose effect is to construe the pronoun in the antecedent in (12a) as a variable bound by a λ -operator attached to the VP, in the way informally illustrated in (12b) (using our notation). The result of such a construal rule is interpreted as in (12c) and gives us the right antecedent for the VP-anaphor in the second conjunct of (12a). Perhaps, the mechanism that I am suggesting might be responsible for de se readings can be regarded as a generalization of Williams's pronoun rule. The idea is the following. Williams's pronoun rule (or some descendant of it) optionally binds a pronoun by means of an operator adjoined to the VP. There seems to be no a priori reason why such an operator could not have scope also at other admissible sites, i.e., most prominently, at a clausal level. This would give us precisely what we need for the de se readings.

As the presence of an operator in the antecedent VP in (12) is associated with the availability of a sloppy reading for the pronouns bound by it, one should expect similar sloppy phenomena to occur with de se pronouns (i.e. pronouns bound by operators adjoined to S'). As we shall see, this expectation is indeed warranted.

What I have said is sketchy and informal. Exploring the way in which the GB binding theory interacts with the present proposal in a broader perspective (and hence its ultimate viability within such a framework) exceeds what I can try to do here.⁷ What I have said should suffice to lend at least some preliminary plausibility to our basic idea.

2.2. A categorial approach

Let us now consider briefly a categorial execution of our Lewis-style approach to de se phenomena. One can introduce in an extended categorial grammar (cf. Bach and Partee (1980)) lexical entries for pronouns of the form shown in (13)

- (13) $\langle \text{she}^*, \text{NP}, x_n, \langle Q_n \rangle \rangle$, where $Q_n = \lambda p \lambda x_n [p]$

The entry in (13) is a tuple composed of a phonological representation of the pronoun, its syntactic category, its logical translation and a series of stores. *-pronouns function like wh-traces in Cooper (1983) and related work in that they come with an operator in store. Such an operator (let us call it "a *-operator") is a propositional abstractor: it forms a property out of a proposition. *-operators are passed up the tree to be retrieved, as usual, at S-level. The relevant retrieval process follows the pattern characteristic of such processes, illustrated in (14):

- (14) a. $S \Rightarrow S$
 $Q_n \in \text{ST} \quad \emptyset$
 $\quad \quad \quad 1 \quad \quad 2$
 b. $2' = Q_n(1')$

Rule (14) enables us to remove a *-operator from store and simultaneously applies it to the meaning of the sentence (i.e. a proposition) turning it into a property. Let us see how this works by means of an example. Rule (14) assign the analysis in (15b) to the sentence in (15a)

- (15) a. Mary believes that she is in danger
 b. $\langle \text{Mary believes that she}^* \text{ is in danger, S, believe}'(M, \lambda x_3 [\text{in danger}'(x_3)]) \rangle$
- $\langle \text{Mary, NP, M} \rangle \quad \langle \text{believes that she. is in danger, VP, } \lambda y [\text{believe}'(y, \lambda x_3 [\text{in danger}'(x_3)])] \rangle$
 $\langle \text{believe, VP/S, believe}' \rangle \quad \langle \text{she. is in danger, S, } \lambda x_3 [\text{in danger}'(x_3)], \emptyset \rangle$
 $\quad \quad \quad \langle \text{she. is in danger, S, in danger}'(x_3), \lambda p \lambda x_3 [p] \rangle$
 $\quad \quad \quad \langle \text{she.}, \text{NP}, x_3, \langle \lambda p \lambda x_3 [p] \rangle \rangle \quad \langle \text{is in danger, VP, in danger} \rangle$

If we have a sentence with two *-operators in store and we try to retrieve

them at the same level, the result will be ill-formed. The structure that we would obtain would have the form $Q_i(Q_j(S'))$. But Q_j will turn S' into a property. Hence Q_i will be unable to apply to it, for *-operators are propositional operators. Rule (14) as currently specified does not prevent retrieval of a *-operator at root level, but it could easily be modified to do so, and I will not try to settle how this should be done here.

An embedded clause containing a *-pronoun is interpreted as a property. This property is understood as being predicated of one of its coarguments, indirectly, via the entailments associated with de se constructions. We must therefore ensure agreement between the argument that acts as controller of the property and the relevant *-pronoun. For example, we must disallow sentences identical with (15) but with a masculine he. This is an analogue of the requirement that the operator introduced in the GB-variant of the present approach must be coindexed with a suitable antecedent. There are various ways of accomplishing this effect. The approach I have in mind is the theory of predication developed in Chierchia (1987a), but any variant of the "Control Agreement Principle" of Gazdar et al. (1985) would also do.

These remarks, sketchy as they may be, seem to give some substance to the claim that a Lewis style approach to de se phenomena can be implemented in semantics in a fairly systematic way. This enables us to test some further consequences that I think the present approach yields, to which we now turn.

3. CONSEQUENCES

3.1. Infinitives and gerunds

Suppose there are other constructions besides that-clauses that on independent grounds we have reasons to believe denote unsaturated structures, i.e. properties. Suppose, furthermore, that some propositional attitude verbs subcategorize for such structures. It follows from our theory that such verbs as they take property-denoting items ought to express unambiguously de se attitudes.

Now, are there other constituents that can be maintained denote properties? I think so: infinitives and gerunds. I have tried to argue in Chierchia (1984) that the validity of arguments such as those in (16) provides evidence that the semantic value of infinitives and gerunds is something like a property (whatever the syntactic category of such constituents may be).

- (16) a. Pavarotti tried/practised/began everything that Domingo tried/practised/began

- (16) b. Domingo tried/practised/began singing Rigoletto
 c. Pavarotti tried/practised/began singing Rigoletto

If *try*, *practise* or *begin* are relations of agents to properties (types of activities), then the validity of the argument in (16) follows in a way that hardly requires any comment.

The same point can be made from a slightly different angle. Consider the sentences in (17).

- (17) a. Domingo practised singing Rigoletto
 b. Pavarotti practised it too
 the thing that Domingo practised
 what Domingo practised
 c. practise (P,x)
 d. practise (P, λx [practise (D,x)])

If gerunds denote properties, all the referential NP's in (17b) are predicted to have *only* a sloppy interpretation. I.e. it is predicted that the sentences in (17b) can only mean that Pavarotti practised his own singing. Simply put, the pronominal in (17b) is presumably interpreted as a variable, so that the interpretation of the relevant sentence is something like (17c). The definite NP's and the free relative are interpreted as definite descriptions, so that the interpretation of the relevant sentence is something like (17d).⁸ In any case, what they can refer to will then depend on what the matrix verb, i.e. *practise*, can take as argument. *Practise* only takes property-denoting items. Thus, (17b) can only express a relation between Pavarotti and a property, which yields an indirect attribution of the property to Pavarotti, along the lines made explicit for belief. Whence the sloppy interpretation of the NP's in (17b) and de se character of the constructions in question.

Similar arguments can be build^t I think, for infinitives. As a matter of fact, we find one such argument very clearly stated in Jerry Fodor's *Language of Thought* (1977: 142ff). We report it in (18):

- (18) a. The cat wanted to eat the cheese
 b. The mouse got what the cat wanted
 c. The mouse got to eat the cheese

Fodor assumes that the logical form of (18b) is roughly as given in (19) and wonders what the logical form of the first premise (18a) should be to get the inference to follow.

- (19) $\exists x$ [the cat wanted $x \wedge \forall y$ [the cat wanted $y \rightarrow x=y$] \wedge the mouse got x]

Clearly, Fodor argues, no standard propositional argument for the complement of *want* in (18a) will do. Suppose for example we assume that (18a) has the logical form in (20).

- (20) the cat (λy [y wants that y eats the cheese])

From (20) and (18) it ought to follow that what the mouse got is that the cat eat the cheese, which is not what we want.

I think that Fodor is right: no sound proof theory will get us from (18a, b) to (18c) if the relatum of *want* in (27a) is a genuine proposition, without additional assumptions.

It is also instructive to consider Fodor's attempt to solve this problem. He takes the facts in (18) show that the "language of thought" has got to be essentially richer than ordinary logic. In particular it has to contain an element, which he refers to as *self*, that acts as the subject of infinitives. He calls such an item a "variable variable" and it somehow has the capacity of licensing the inference in question.

In current theories, the "variable variable" *self* has been replaced by PRO, the null pronominal element that various arguments suggest acts as the syntactic subject of infinitives and gerunds. But the replacement of *self* by PRO does not make the problem posed by (16) and (18) any easier: why are these inferences valid? How are we to interpret infinitives and gerunds? It seems that if we interpret them as genuine propositional creatures, we would deprive ourselves of what appears to be the most straightforward way to understand the validity of the inferences in question (but see the discussion below towards the end of this section).

Suppose, on the other hand, that PRO headed sentences are regarded as unsaturated structures. I.e. suppose that PRO is interpreted as, say, a λ -abstractor and that, consequently, infinitives and gerunds are analyzed as properties. Assume, in other terms, that a structure such the one in (21a) is semantically analyzed as in (21b):

- (21) a. PRO to eat cheese
 b. λx [x eats cheese]

Then the validity of (16) and (18) becomes straightforward: it follows from a fully compositional semantics for the sentences involved and elementary logic. A puzzling series of facts, falls into place with a rather minimal shift of perspective. The "variable-variable" of Fodor's is just an abstracted variable.

Perhaps, the interpretation of PRO as a property-abstractor is also mediated by an operator at LF, so that the structure of, e.g. (22a) would be as in (22b).

- (22) a. The cat wants to eat the cheese
b. the cat_i wants O_i[PRO_i to eat the cheese]

This line of analysis is essentially a variant of Williams's view of control. Williams regards control as a relation between an antecedent and a sentence, construed as a derived predicate. The fact that infinitives and gerunds behave semantically as properties supports, I think, such a view.

Notice that, if our argument is correct and infinitives and gerunds are to be interpreted as properties, it follows that PRO *must* be associated with an abstractor, while the presence of an abstractor to bind pronominals like *s/he* is optional, as we have seen. One might wonder whether there are phonologically realized counterparts of PRO, i.e. pronominal elements obligatorily associated with an abstractor. I think that there are. They will be discussed in sec. 3.4.

Now, if infinitives denote properties and infinitival-taking verbs relations of agents to properties, how come we usually understand infinitives as if they had a subject? For example, *John wants to be fired* means, roughly, that he wants to bring about a situation where he (and not somebody else) is fired. *John* controls the complement of *want* in the example at hand. The answer that suggests itself is quite obvious: the relations at hand, like the believe-relation are "self-ascriptive". They involve an indirect attribution of the property in question to the relevant argument.⁹ For example, (23a) entails (23b).

- (23) a. Pavarotti wants very much to get help
b. Pavarotti wants very much for Pavarotti to get help

The semantic aspect of control can be simply our (implicit) knowledge of such entailment, which comes with knowing the lexical meaning of *want*.

Notice, however, that (23b) does not entail (23a). The situations that show this are precisely those that are involved in *de se* phenomena. The asymmetry of the entailment relation in (23) is just what we found in (6), sec. 1.2. This explains why PRO, the subject of infinitives, will in general be interpreted *de se*, and unambiguously so. Such an observation has occasionally been made (e.g. by Partee (1975) or by Chomsky in class lectures) but never pursued.

To sum up, we have argued, following Lewis, that *de se* phenomena involve relations to properties (that entail relations to propositions). On independent grounds (provided by the validity of the arguments in (16)

and (18)) we have seen that there are good reasons to believe that infinitives and gerunds denote properties. It follows that if relations involving infinitives yield an (asymmetric) propositional entailment, they will unambiguously characterize attitudes *de se*. The facts seem to bear this prediction out.

Nothing excludes, it should be noted, that some relations involving infinitives entail and are also entailed by the relevant corresponding *de re* propositional constructions. In such a case, *de se* and *de re* readings will collapse into one. Compare, for example the sentences in (24) with those in (25):

- (24) a. John persuaded Mary to be fired
b. John persuaded Mary to bring about a situation where she is fired
(25) a. John forced Mary to leave
b. John forced Mary to bring about a situation where she leaves
c. make, succeed in, be nice of, etc.

Sentence (24a) asymmetrically entails (24b). Sentence (25a), on the other hand entails and is entailed by sentence (25b). And (24a) has unambiguously a *de se* interpretation, while (25a) has unambiguously a *de re* interpretation, exactly how our theory would predict. (Causative or evaluative predicates, such as those listed in (25c) seem to behave just like *force*).

The contrast in (24)-(25) is subtle but quite clear, and it seems to me provides strong evidence against any attempt to assimilate *de se* readings with specific items at logical form (be it PRO or something else). The constructions in (24) and (25) are identical as far as syntax and logical form goes. The explanatory force of our theory in this connection stems, I believe, from the fact that *de se* readings fall out of two hypotheses that can be tested independently: the fact that certain constituents denote properties and the entailments associated with the relevant constructions. Having a relation of individuals to properties is necessary but not sufficient to license attitudes *de se*.

I should point out an open problem in connection with the present approach, concerning the interpretation of reflexives in exceptional case marking (or raising-to-object) constructions and *for-to* clauses, illustrated in (26).

- (26) a. John wants very much for himself to get help
b. John believes himself to be in danger
c. John expects himself to get help

Given standard assumptions about the meaning of these constructions,

our theory would predict it to be ambiguous between a *de se* and a *de re* reading. But it seems to me that at least with (26a,b) the purely *de re* reading is not salient, if available at all. Our grammar could, of course, force the *de se* interpretation here. But it would be nicer to link this fact to some other peculiarity of the constructions in question (such as, perhaps, the exceptional way in which the presence of a reflexive is licensed here). Things are made even worse by the fact that a purely *de re* interpretation seems to be available for (26c), especially if it is contrasted with a sentence like *John expects to get help*. At present, I have no non ad hoc account for the facts in (26).

Finally, it should be noticed that talking of self-ascriptive relations in the case of object control verbs like *persuade* is perhaps misleading. The subject of the attitude is not the one who does the self-ascription. We will nevertheless keep using this label. The definition of "self-ascriptivity" can be easily generalized to every relevant relation, as illustrated in (27).

- (27) an n-place relation R is "self-ascriptive" iff it (asymmetrically) licenses the following entailment:
 $R(\dots x \dots Q \dots) \rightarrow R(\dots x \dots Q(x) \dots)$

Self-ascriptivity is simply the semantic aspect of what linguists call control.

So PRO (i.e. infinitives and gerunds) is one of the ways in which language singles out *de se* relations. Hence, *de se* relations are, if not metaphysically special, semantically special. Furthermore, the facts discussed in this section, among other things, provide a strong argument against a propositional semantics for *de se* reports, such as, for example Cresswell (1985). The point is that infinitives ought to be treated *de se*, as that appears to be their standard meaning. If an attitude *de se* is a relation to a propositional creature, say a set of speaker-centered worlds, as for Cresswell, then the validity of the arguments in (16) and (18) and the sloppy character of the anaphoric NP's in (17) becomes hard to explain.

To see why, it might be instructive to pursue what moves one might want to make in order to maintain a propositional approach to *de se* phenomena. To this task I will devote the remainder of this section.

One could maintain, for example, that the interpretation of the NP's in (17b), reproduced below, involves replacing them with a copy of their antecedent, as sketched in (28b).

- (28) a. Pavarotti practised it too
 the thing that Doming practised
 what Domingo practised
 b. Pavarotti practised [PRO_i singing Rigoletto]
 c. *Pavarotti_i practised [PRO_i VP]

If the index in (28b) happened to be different from the one of the matrix subject, the resulting structure would be ruled by whatever rules out (28c) in the first place.

A priori, this strategy might have some plausibility for pronouns (but see below section 3.3 for some problems). However, definite descriptions and free relative do not seem to normally involve copying operations of this kind. Consider, for example (29):

- (29) A man_i was talking to a woman_j. The man_j was wearing a hat.

Suppose we try to replace *the man* in (29) with a copy of its antecedent. If we pick a *man_i*, we would have to somehow ensure that the man we are talking about in the two sentences of (29) is the same.¹⁰ In other terms, it remains to be seen how this line fits with a general theory of definite descriptions, a task that *prima facie* appears to be non trivial.

Alternatively, one might try to maintain that the interpretation of the NP's in (28a) calls for some kind of accommodation, in the sense of Heim (1982). What I have in mind can be illustrated by means of the following example (which I got from J. Higginbotham):

- (30) a. John ate a ham sandwich
 b. Mary ate what John ate
 c. Mary ate a ham sandwich

The idea is that if we try to interpret the free relative in (30b) literally as the thing that John ate, we get a pragmatically deviant reading. This forces us to accommodate by interpreting (30b) as, say, "Mary ate something of the same kind as what John ate". Similarly, if we try to interpret "what Domingo practised" literally in (18a), we get a pragmatically deviant sentence, which induces accommodation.

While this strategy too is a priori plausible, the problem that I see with it is that it doesn't seem to work for pronouns. Contrast (31a-b) with (31c-d).

- (31) a. John ate a ham sandwich
 b. Bill ate it too
 c. Domingo practised singing Rigoletto
 d. Pavarotti practised it too

In (31b) we do get a contradictory reading. (31a-b) taken as a discourse suggests that John and Bill ate the same sandwich. But this seems to show that the accommodation strategy is not readily available here, after all. The

discourse (31c-d), on the other hand is perfect. Why should these two contexts behave differently, in spite of their complete parallelism?

This does not show that propositional theories of *de se* phenomena are doomed to failure. But the propositional views that more readily come to mind do appear to be beset by non trivial difficulties. In contrast, if we assume that the semantics of *de se* is mediated by properties, the phenomena under consideration fall rather smoothly into place.

3.2. Subcategorization

There is another aspect of our approach to *de se* reports perhaps deserves to be discussed. Take a verb that expresses a mental attitude such as, say *believe* (or *know*). A *de se* belief is bearing the believe-relation to a property. Infinitives (and gerunds) are in languages like English ways of referring to properties. Why doesn't, then, English allow *believe* (or *know*) to subcategorize for infinitives? Isn't this a curious circumstance?

Our theory indeed leads us to expect this circumstance to be odd, i.e. an idiosyncrasy of English. If the relation between syntax and semantics is reasonably transparent (however such transparency is captured), there should be a systematic correspondence between the type of semantic objects that a relation can take as arguments and the syntactic category of the complements of the corresponding verb. In absence of evidence to the contrary, it is useful to maintain that there aren't wild divergences on this score.

In the particular case at hand, mental attitudes are relations that admit semantically properties as one of their relata. Hence, one should expect that any verb that expresses some such relation could, in principle, subcategorize syntactically those constituents that denote properties.

This expectation is in fact warranted in many languages. Italian is an example, as illustrated in (32):

- (32) a. Pavarotti crede di essere in pericolo
Pavarotti believes to be in danger
b. Pavarotti sa di essere in pericolo
Pavarotti knows to be in danger

The sentence in (32) expresses unambiguously a *de se* reading, as our theory would predict. Another case in point is *seem* as illustrated in (33)

- (33) a. it seems to Pavarotti that he is in danger
b. a Pavarotti sembra di essere in pericolo
(it) to Pavarotti seems to be in danger

This case is fully parallel to the *believe* case.

The point can be made in the following terms. Here is the typology that we expect on the basis of the present approach. First, we might have 1-place predicates of propositions (like *be likely*). These predicates will not be able to express attitudes *de se*, simply because they are not attitudes of agents to propositions in the first place. Hence, these predicates will not alternate S-complements with property denoting complements such as infinitives. Whence the ungrammaticality of, e.g. *it is likely to love Mary* (vs. *it is hard (for John) to love Mary*). Second, there are predicates which express relations of agents to propositions. These are all also capable of having properties as arguments, since any propositional attitude can also be an attitude *de se*. Hence, in general, we expect predicates of this second type to be able to subcategorize for property-denoting items (such as infinitives and gerunds), modulo accidental gaps. Third there are predicates which express purely *de se* relations of agents to properties. These predicates will only subcategorize property-denoting items. The present typology is summarized in (34).

- (34) a. propositional predicates: only S (example: is likely)
b. relations of individuals to properties and propositions: tensed S and INF/GER (example: believe, know, etc.)
c. relations of individual to properties: (subjectless) INF/GER (example: try, practise, etc.)

The above typology embodies a non trivial claim on what relations can be lexicalized in the languages of the world. It claims that relations of individuals to propositions should alternate with relations to properties, which has consequences for syntactic subcategorization: it provides us with a markedness scale as to what to expect.

This is of course only a rough first approximation. But the Italian facts appear to provide some preliminary support for it.

3.3. Anaphora with proposition-taking verbs

The present theory has consequences also in a different area, namely the anaphoric behavior of *that*-clauses. *That*-clauses can be interpreted either as propositions or (if they contain a pronoun) as unsaturated structures (i.e. properties). Consider now the discourse in (35a-b)

- (35) a. Domingo believes that he is a genius
b. Pavarotti believes it too
the thing that Domingo believes
what Domingo believes

- 35 d. believe (P, x)
e. believe (P, $\lambda x[\text{believe}(D,x)]$)

The NP's in (35b) are interpreted just like the ones in (28a) above, namely as shown roughly in (35c-d). But now, *believe* can take as arguments both properties and propositions. Hence, we should expect both a strict reading for these NP's (in case they are anaphorically linked to a proposition) and a sloppy one (in case they are linked to a property) depending on how the antecedent is disambiguated. And in fact, this seems exactly right, as shown by the fact that we can continue the discourse in (35) as in (36a) or (36b), on the basis of how (35a) is understood.¹¹

- (36) a. Pavarotti believes that he (Pavarotti) is a genius
b. Pavarotti believes that he (Domingo) is a genius

The sloppy reading of the NP's in (35b) is the one linked to the *de se* interpretation of the embedded pronoun. Thus our theory predicts a systematic contrast between the constructions in (35a) and the ones in (28), contrast that seems to be borne out.

It should be remarked that the availability of a sloppy reading disappears if the antecedent to (35b) is taken to be a pronounless structure like the one in (37):

- (37) a. Domingo believes that Domingo is a genius
b. Domingo believes that the author of this book is a genius

Sentences of this kinds are slightly odd due to a principle C effect, i.e. the fact that a referential NP usually is not taken to corefer with another referential NP C-commanding it. However, it has been noted that these sentences can be rescued in suitably contrastive contexts (see Reinhart (1983) for relevant discussion). Yet, even when they are grammatical, such sentences never support sloppy anaphora, as antecedents of NP's such as those in (35b). This of course, follows on the present theory, as the sentences in question cannot have a *de se* interpretation (i.e. be interpreted as properties). The latter interpretation is specifically linked to the presence of a pronoun that we can abstract over. If however one takes *de se* interpretations to be just a case of *de re* interpretations (as suggested by Boer and Lycan (1980)), there would be no reason why (37a-b) should lack a *de se* interpretation. And we would be left without an account for the above facts.

These considerations suggest yet another way in which attitudes *de se* are semantically special: they are associated to that-clauses containing

pronominal elements, but not in that-clauses containing non-pronominal, referential NP's.

The conclusions suggested by the facts in (35) are further supported by inferences such as those in (38) and (39)

- (38) a. Domingo believes that he is a genius
b. Pavarotti believes everything that/whatever Domingo believes
c. Pavarotti believes that he (Pavarotti) is a genius
d. Pavarotti believes that Domingo is a genius
(39) a. Pavarotti believes that he (Pavarotti) is a genius
b. Domingo believes that he (Domingo) is a genius
c. There is something that both Pavarotti and Domingo believe

This contrasts systematically with the inferences we are able to draw with try-type verbs, illustrated in (16) above. Try-type verbs subcategorize only property-denoting structures and hence only one pattern of inference is possible. Here however, two patterns of inference become possible, depending on how the first premise is construed. The sloppy inference in (38) is associated with the *de se* reading of the embedded pronoun.

Similarly, if (39a) and (39b) are interpreted as attitudes *de se*, then both Pavarotti and Domingo will be disposed to assent to the sentence "I am a genius" as uttered by them, and hence they do have a common belief.

Essentially the same contrast shows up with respect to VP-anaphora. Consider the examples in (40) (due to T. Reinhart):

- (40) a. John wants to become a doctor
b. but his mother doesn't want to ____
c. but his mother doesn't want that

The sentence in (40b) has only a sloppy reading (meaning "his mother doesn't want to become a doctor"). This follows from any current approach to VP-anaphora that I am familiar with. On the other hand, (40c) has both a strict and a sloppy reading, as our theory predicts. *Want* subcategorizes for both PRO-headed infinitives and infinitives with an overt lexical subject. Hence it is both property- and proposition-taking and the semantic value of the pronominal *that* can be construed accordingly.¹²

Suppose, now, we were to say that interpretation of the pronominal *that* in (39c) goes through a copying operation, such as the one sketched at the end of sec. 3.1. It would seem that such a copying operation could yield only the structure given in (41), just as was argued for (28c) above:

- (41) his mother_i doesn't want [PRO_i to become a doctor]

This gives us the sloppy reading. How is then the strict reading to be obtained? A separate mechanism seems to be called for. And it remains to be seen what blocks this mechanism, whatever it may be, from being operative in the case of try-type verbs, where only sloppy readings are available. This casts further doubts, I believe, on propositional theories.

3.4. Long distance reflexives

It has been noted by various people that long-distance reflexives are often "self-oriented". For example, Sells (1987), building on previous work on logophoricity, argues that the interpretation of long-distance reflexives is mediated by certain discourse roles. One of these roles is what he calls *self*. Such a role is associated with the argument of a relation whose mental state or attitude is being described. Sells adopts a Kamp-style discourse representation theory where roles such as *self* are overtly marked. A pronoun is linked to its antecedent via such a role marker.¹³

In the terms of the present work, Sells's *self*-oriented long-distance reflexives appear to constitute a case of pronominal elements associated with a *de se* interpretation. As an illustration, consider the following example:

- (42) a. Pavarotti crede che i propri pantaloni siano in fiamme
Pavarotti believes that self pants are on fire
b. Pavarotti crede che i suoi pantaloni siano in fiamme
Pavarotti believes that his pants are on fire

The non-reflexive pronoun in (42b), just like its English counterpart can have both a *de se* and a *de re* interpretation. However, the reflexive pronoun in (42a) appears to have only a *de se* interpretation. (42a) would be false, I believe, in the mirror-situation described in sec. 1.1.

This can be further seen by the fact that (43a), an example I got from Andrea Bonomi, is not contradictory, while (43b) is. (43b) parallels the contradictoriness of (43c) (where '#' marks a contradiction).

- (43) a. Pavarotti crede che i suoi pantaloni siano in fiamme. Ma non si e' accorto che i pantaloni sono i propri.
Pavarotti believes that his pants are on fire. But he hasn't realized that the pants are his own
b. # Pavarotti crede che i propri pantaloni siano in fiamme. Ma non si e' accorto che i pantaloni sono i propri.
Pavarotti believes that self pants are on fire. But he hasn't realized that the pants are his own

- (43) c. # Pavarotti crede di avere i pantaloni in fiamme. Ma non si e' accorto che i pantaloni sono i propri
Pavarotti believes [PRO to have burning pants]. But he hasn't realized that the pants are his own

Thus *proprio* in Italian patterns just like PRO and it provides us with another way in which languages single out relations *de se*. Perhaps, *proprio* might be viewed as a phonologically realized counterpart of PRO.

Given the theory that we have developed, we are forced to assume that the link between *proprio* and its antecedent must be mediated by an operator adjoined to the relevant S'. So for example (42a) must have the structure in (44a), which yields the interpretation in (44b).

- (44) a. Pavarotti_i crede O_j [che i propri_i pantaloni sono in fiamme]
b. believe (P, $\lambda x[x$'s pants are on fire])

The formula in (44b) expresses the by now familiar self-attributive belief-relation between Pavarotti and the property of having burning pants.

So, as a first shot, one might say that long-distance reflexives in Italian *must* be operator-bound. This operator is adjoined to S' or (if no S'-site is available¹⁴) to VP. If the operator is adjoined to VP, as in VP-ellipsis cases, the resulting abstract is predicated of the subject of the VP. Otherwise, it is (indirectly) predicated of a suitably prominent coargument, namely the one whose attitudes are being described.¹⁵ I would maintain that a process of this sort is what is involved in the truth-conditional interpretation of what Sells calls *self*-oriented long-distance reflexives.

There are two consequences that, in spite of the sketchiness of the present proposal, we are able to draw from it. The first concerns the non local character of the reflexive-antecedent relation in the case of *proprio*. As, in general, the relation between the abstractor (that we represent as O) and the pronoun abstracted over is not subject to island constraints, the relation between a long-distance reflexive and its antecedent shouldn't either. This appears to be borne out, and appears to confirm our hypothesis that we are not dealing with a movement dependency:

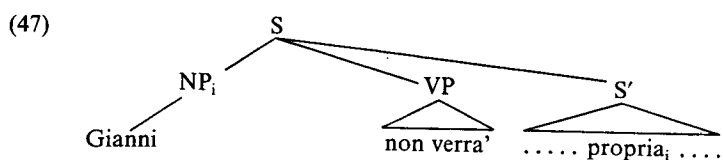
- (45) Pavarotti_i crede O_j [che il fatto che i propri_i dischi abbiano tanto successo sia un miracolo]
Pavarotti believes that the fact that self records are so successful is a miracle

The second consequence is perhaps more interesting. There is a well-known constraint that limits the range of possible antecedents for *proprio*. A reflexive embedded in an adverbial clause in general cannot have as

antecedent the subject of the sentence immediately containing it. This is illustrated in (46).

- (46) a. *Gianni_i non verri' a meno che non inviti la propria_i moglie
Gianni will not come unless (I) invite self wife
b. Gianni_i pensa che sia bene non venire a meno che non inviti la propria_i moglie
Gianni thinks that it would be good not to come unless (I) invite self wife

I assume that the structure of (46a) is roughly as indicated in (47)



At a descriptive level, one can say that a long-distance reflexive of this kind must be embedded in a subcategorized position with respect to its antecedent. What is interesting is that this constraint does not appear to be an idiosyncrasy of Italian, but seems to hold of several other unrelated languages, such as, e.g. Icelandic (see Sells (1987) and references therein). It thus seems to be a non accidental property of the kind of binding involved in long-distance reflexive systems.

Notice that there is nothing wrong with the binding relation in (47) as such (whether one assumes an intervening operator or not): the antecedent C-commands the pronoun and in fact binding of (non reflexive) pronouns by quantified NP's is acceptable in such a configuration, as (48) illustrates:

- (48) Nessuno studente_i verri' a meno che non lo_i inviti personalmente
No student will come unless (you) invite him personally

So the question is of course: why? Where does this restriction come from? I am not aware of any hypothesis with great explanatory force in this connection. In general, the theories I am familiar with (such as Giorgi's) simply set up a notion of domain that excludes S-adverbials from the domain of their immediate subject. The present theory, on the other hand, might enable us to actually derive this domain restriction. I will now try to state what the idea is in this connection in a way which is, as far as possible, theory neutral.

A priori, looking at (47) there are four possible sites for the operator in question: the adverbial S', the root S', the adverbial VP, or the root VP.

It turns out that none of them will do. Basically, if we try to attach the operator to the matrix S', it would be without antecedent. Differently put, the structure would be interpreted as property rather than as a proposition, and hence uninterpretable as an ordinary declarative root-sentence. If the operator in question is attached to the matrix VP it would not be able to bind the reflexive, as the latter would not be in its scope. If the operator is attached to the adverbial's VP, its antecedent would have to be the subject internal to the adverbial clause. Such structures are indeed well-formed but of course lack the relevant reading (Cf. *Gianni non verri' a meno che Mario_i O_i[_{VP}non inviti la propria_i moglie]*, "Gianni will not come unless Mario_i invites self_i wife"). Finally, if we try to adjoin our abstractor to the adverbial S', we get a property in the wrong position. Such a property is neither in VP-position, the canonical predicative position, nor is the argument of some higher function that can give rise to an indirect attribution to some other argument. Consequently, the resulting structure is uninterpretable.

This hypothesis calls of course for further testing, which presumably would require making more explicit assumptions about the structure of S' (and of COMP) than what we are able to do here. Be that as it may, the point is that on our approach long distance reflexives must be interpreted either by means of a VP-level operator that ensures binding by the subject (see e.g. Bach and Partee (1980)) or by means of an S'-level operator that ensures a de se interpretation. In the latter case, however, the resulting property must occur as argument of a (self-ascriptive) propositional-attitude relation. If we abstract over the adverbial S' in (48) neither option is actualized, whence the ungrammaticality of the resulting structure.

I am aware of the sketchiness of the present proposal vis-a-vis the complexity of the distribution of long distance reflexives. Furthermore the suggested parallelism between *proprio* and PRO needs to be more carefully weighed. However, long-distance reflexives could not go unmentioned, as they do appear to be systematically linked to a de se interpretation. Our theory of de se forces us to analyze them as bound by means of a property-abstractor. What I hope to have shown here is that this idea is worth pursuing.

4. CONCLUSIONS

There are many questions, some foundational other empirical, that I had to leave open in the preceding discussion. However, let me try to summarize what I believe comes out of it.

There are many conceivable cognitive and causal relations that link us to the objects of our mental attitudes, in a de re mode. Perhaps, the cognitive

access that we have to ourself is naturally classified among other causal and cognitive *de re* links. Or perhaps it is a more fundamental type of cognitive access. Be that as it may, such a cognitive access is semantically special, in the sense that it is singled out in a number of ways by the semantic system of natural language. It is systematically excluded from the interpretation of (non-pronominal) referential expressions. It is systematically present in the interpretation of overt pronouns. It is systematically and unambiguously associated with the interpretation of PRO the null subject of infinitives and gerunds. It is associated with the interpretation of long-distance reflexives (at least in some languages).

Our approach is based on the following idea. Attitudes *de se* are represented as relations to properties. A relation to a property has to be self-ascriptive to be understood *de se*. Intuitively this means that the property has to be indirectly predicated of the bearer of the attitude. Formally, self-ascriptivity can either be taken as a primitive, as Lewis does, or explicitly defined in terms of the entailment patterns associated with the relation, as we have tried to do here.

To turn this into a substantive semantic claim, all we need to assume is that we refer to properties perhaps more extensively than one would have suspected. Notice that I haven't said what I take properties (or, for that matter, propositions) to be, even though I do have strong feelings on these issues. For our present purposes, all that matters is that properties are "subjectless" incomplete structures whose completion results in a proposition. Our key assumptions are schematically summarized in (49)

- (49) a. PRO headed structures, i.e. infinitives and gerunds, (and long distance reflexives) are systematically interpreted as derived predicates or as open, unstaturated structures (i.e. as properties)
 b. that-clauses with a pronoun can be interpreted as derived predicates

These claims are *per se* very simple and seem to be at least compatible with widely shared assumptions concerning the syntax of the relevant constructions. From (49) we are able to derive the following facts:

- (50) a. lack of *de se* interpretations for non pronominal NP's (cf. (37))
 b. availability of *de se* interpretation for embedded overt pronouns
 c. the distribution of *de se* readings in PRO-headed structures (i.e. the contrasts in (24)-(25))
 d. the pattern of validity for inferences involving infinitives and gerunds (i.e. (16) and (18))
 e. the pattern of validity for inferences involving that-clauses (i.e. (38)-(39))

- (50) f. the contrast in the interpretation of anaphoric NP's with try-type vs. believe-type verbs (i.e. (17) vs. (35))
 g. properties of subcategorization (i.e. (34))
 h. interpretive constraints on the long-distance antecedent-anaphor relation (i.e. (46))

If (49) is correct, it is not easy to see how the facts that I have just listed could actually be otherwise. At present I am not aware of an equally principled way in which such facts could be derived from a theory that treats verbs of attitudes as relations involving solely propositions (whatever one takes propositions to be).

Minimally, this shows that the hypothesis that the cognitive access to self is encoded as a relation to properties in the semantic system of English is non trivial. More generally, it shows that a theory of *de se* belief can advantageously be made responsive to the kind of evidence that generative linguists have been sensitive to and can have a perhaps far reaching impact on understanding how language works.

NOTES

- For example, some authors (like Castañeda (1966), Perry (1979) and Lewis (1979)) argue that attitudes towards oneself cannot be analyzed just as attitudes towards propositions (or propositional creatures). Others (e.g., Boer and Lycan (1980), Stalnaker (1981)) have argued that it is possible to reconcile the special perspectival phenomena that attitudes towards oneself give rise to with the view that mental attitudes are propositional in character.
- In the present paper we shall not discuss *de dicto* beliefs.
- This is not what Lewis proposes, although, as far as I can see it is compatible with it.
- There is a parallelism here with sentences like the one in (1a):
 (1) a. everyone likes someone
 b. $\forall x \exists y [\text{like}(x,y)]$
 c. $\exists y \forall x [\text{like}(x,y)]$

It is generally maintained that (1a) is ambiguous. Its ambiguity is blamed on the different scope that the quantifiers associated with the NP's in (1a) can have. However, the two readings of (1a), represented in (1b-c) are not independent of one another, since (1c) entails (1b).

- See Lewis (1979, sec. XIII) for a proposal along this line. See also Cresswell and von Stechow (1982).
- See on this Lewis (1979, sec. X) and Cresswell and von Stechow (1982). Stalnaker (1981) proposes to use propositional concepts rather than speaker-centered worlds to deal with belief *de se*. That too yields an enrichment of the notion of proposition that gives us enough structure to characterize belief *de se*. I don't know whether there is a formal relationship between Stalnaker's proposal and the other two (i.e. Cresswell's and Lewis's)
- I believe that Aoun and Clark (1986) is very relevant in this connection.
- The argument goes through, I believe, also if we adopt a theory of definite NP's along the lines of Heim (1982).

9. This is a standard assumption in large part of the (early) Montague tradition. See, e.g. Dowty (1985) and Chierchia (1987b) for relevant arguments. See also Higginbotham (1986) for a critical discussion of this view.
10. This can be done by adopting a Heim style theory. However, Heim's approach involves manipulation of variables, whose interpretation is the standard one, while the solution sketched in the text crucially relies on the actual copying of LF-structures (and on well-formedness conditions on LF-structures). There might be a way of putting these two things together in a coherent fashion, but at present I do not quite see how to do it. Certainly I do not see how to do it in a way that involves fewer stipulations than those needed in the proposal I am advocating.
11. I derive these facts from a number of people: (in chronological order) B. Schein, R. Larson, E. Williams. Their relevance for a theory of attitudes de se became clear to me only recently.
12. Ladusaw (1987) points out that the theory advocated here, in its current form might overgenerate, i.e. predict the availability of unattested anaphoric links. The crucial examples he considers involve shifting from tensed to tenseless clauses (or viceversa). I have nothing substantial to say about such cases. So far, my theory makes the right predictions if one limits oneself to anaphora involving only tenseless or only tensed clauses.
13. Sells doesn't provide, however, explicit embedding conditions for the discourse representation structures he posits.
14. As in:
- (a) Gianni_i O_i[V_{PSI} e' meravigliato del fatto che i propri_i genitori siano stati arrestati]
Gianni was struck by the fact that self parents were arrested
15. Although many aspects of my proposal remain to be worked out, one can perhaps envisage how some of the current theories, such as, for example, Giorgi (1984) could be modified along the lines sketched in the text.

REFERENCES

- Aoun, J. and R. Clark (1986), "On Non-overt Operators" ms. USC and UCLA, Los Angeles, Ca.
- Bach, E. and B. Partee (1980), "Anaphora and Semantic Structure" in K.J. Kreiman and A.E. Ojeda (eds.) *Papers from the Parasession on Pronouns and Anaphora*, Chicago Linguistics Society, 10.
- Boer, S. and W. Lycan (1980), "Who Me?" *The Philosophical Review*, 89: 427-466.
- Castañeda, H.-N. (1966) "'He': A Study in the Logic of Self-Consciousness" *Ratio*, 7:130-157.
- Chierchia, G. (1984), *Topics in the Syntax and Semantics of Infinitives and Gerunds*, Ph.D. Diss., University of Massachusetts, Amherst.
- Chierchia, G. (1987a), "Aspects of a Categorical theory of Binding" in R. T. Oehrle, E. Bach and D. Wheeler (eds.) *Categorical grammars and Natural Language Structures*, D. Reidel, Dordrecht, 125-151.
- Chierchia, G. (1987b), "Structured Meanings, Thematic Roles and Control" forthcoming in G. Chierchia, B. Partee and R. Turner (eds.) *Properties, Types and Meaning*, vol. II, D.Reidel, Dordrecht.
- Cooper, R. (1983), *Quantification and Syntactic Theory*, D. Reidel, Dordrecht.
- Cresswell, M. J. (1985), *Structured Meanings*, MIT Press, Boston
- Cresswell, M.J. and A. von Stechow (1982), "De Re Belief Generalized" *Linguistics and Philosophy*, 5: 503-535.

- Dowty, D. (1985), "On Some Recent Analyses of the Semantics of Control" *Linguistics and Philosophy*, 8: 1-41.
- Fodor, J. (1975), *The Language of Thought*, Harsocks, Harvester Press.
- Gazdar, G., E. Klein, G. Pullum and I. Sag (1985), *Generalized Phrase Structure Grammar*, Harvard University Press, Cambridge, Mass.
- Giorgi, A. (1984), "Towards a Theory of Long-Distance Anaphors: A GB-Approach" *The Linguistic Review*, 3: 307-361.
- Heim, I. (1982), *The Semantics of Definite and Indefinite NPs*, Ph.D. Diss., University of Massachusetts, Amherst.
- Higginbotham, J. (1986), "Elucidations of Meaning" ms., MIT, Cambridge Mass.
- Ladusaw, W. (1987), "Inference Patterns from Infinitival Complements" *Proceedings of WCCFL 6*, Stanford, Ca.
- Lewis, D. (1979), "Attitudes De Dicto and De Se" *The Philosophical Review*, 88: 513-543.
- Partee, B. (1975), "Deletion and Variable Binding" in E. Keenan (ed.) *Formal Semantics of Natural Language*, Cambridge University Press, Cambridge, U.K., 16-33.
- Perry, J. (1979), "The Problem of Essential Indexicals" *Nous*, 13: 3-21.
- Reinhart, T. (1983), "Coreference and Bound Anaphora: A Restatement of the Anaphora Question" *Linguistics and Philosophy*, 6: 47-88.
- Sells, P. (1987), "Aspects of Logophoricity" *Linguistic Inquiry*, 18: 445-479.
- Stalnaker, R. (1981), "Indexical Belief" *Synthese*, 49: 129-151.
- Stalnaker, R. (1984), *Inquiry*, MIT Press
- Williams, E. (1977), "Discourse and Logical Form" *Linguistic Inquiry*, 8: 101-139.

Groningen-Amsterdam Studies in Semantics (GRASS)

This series of books on the semantics of natural language contains collections of original research on selected topics as well as monographs in this area. Contributions from linguists, philosophers, logicians, computer-scientists and cognitive psychologists are brought together to promote interdisciplinary and international research.

Editors

Alice ter Meulen
Martin Stokhof

Johan van Benthem
University of Amsterdam

Henk Verkuyl
University of Utrecht

Editorial Board

Renate Bartsch
University of Amsterdam

Co Vet
University of Groningen

Other books in this series:

1. Alice G.B. ter Meulen (ed.)
Studies in Modeltheoretic Semantics
2. Jeroen Groenendijk, Theo M.V. Janssen and Martin Stokhof (eds.)
Truth, Interpretation and Information
3. Fred Landman and Frank Veltman (eds.)
Varieties of Formal Semantics
4. Johan van Benthem and Alice G.B. ter Meulen (eds.)
Generalized Quantifiers in Natural Languages
5. Vincenzo Lo Cascio and Co Vet (eds.)
Temporal Structure in Sentence and Discourse
6. Fred Landman
Towards a Theory of Information
7. Jeroen Groenendijk, Dick de Jongh, Martin Stokhof (eds.)
Foundations of Pragmatics and Lexical Semantics
8. Jeroen Groenendijk, Dick de Jongh, Martin Stokhof (eds.)
Studies in Discourse Representations Theory
9. Michael Moortgat
Categorial Investigations
10. Irena Bellert
Feature System for Quantification Structures in Natural Languages

All communications to the editors can be sent to:

Department of Philosophy or Department of Linguistics, GN 40
University of Amsterdam University of Washington
Nieuwe Doelenstraat 15 Seattle, Washington 98195
1012 CP Amsterdam U.S.A.
The Netherlands

R. Bartsch, J. van Benthem and
P. van Emde Boas (eds.)

Semantics and Contextual Expression



1989
FORIS PUBLICATIONS
Dordrecht - Holland/Providence RI - U.S.A.