Coreference De Jure

François Recanati

Coreference *de jure*

François Recanati
Institut Jean-Nicod

1. The phenomenon

Two singular terms are coreferential whenever they refer to the same object. Coreference is *de facto* when the two terms merely *happen* to refer to the same object. Sometimes, however, coreference seems to be pre-determined, and arguably guaranteed, in an *a priori* manner. This stronger form of coreference has been given several names in the literature, e.g. 'presupposed coreference' (Fauconnier 1974: 7-8), 'grammatically determined coreference' (Fiengo and May 1996: 122, 2006 : 37), 'explicit coreference' (Taylor 2003), 'strict coreference' (Fine 2007), 'internal coreference' (Lawlor 2010), and 'coreference *de jure*' (Neale 2005, Pinillos 2011, Recanati 2012, Goodsell 2014).[1] The phenomenon is well-known, but there is disagreement regarding its proper analysis.

Fine provides the following criterion for (what I will continue to call) coreference *de jure* :

> A good test of when an object is represented as the same[2] is in terms of whether one might sensibly raise the question of whether it is the same. An object is represented as the same in a piece of discourse only if *no one who understands the discourse can sensibly raise the question of whether it is the same*. (Fine 2007 : 40 ; emphasis mine).

The paradigm case is anaphora. A pronoun and its referential antecedent are coreferential *de jure* : there is no way in which the anaphoric pronoun might not refer to the same thing as the antecedent (assuming the antecedent itself refers). This is guaranteed linguistically, so Fine's criterion applies : whoever has fully understood the statement cannot doubt that there is coreference between the pronoun and its referential antecedent.[3] Coreference, in such cases, is a matter of meaning.

But anaphora is only a special case. There is coreference *de jure* also between two occurrences of the same name-type. As Fine puts it,

---

[1] Other appellations include 'intended coreference' (Kamp 1990: 48), 'presumed coreference' (Lawlor 2001), 'coco-reference' (Perry 2012 : 172), and 'assumed coreference' (Gibbard 2012 : 269-70). But it is not obvious that there is a single phenomenon at stake, rather than several. As we shall see, Kit Fine draws a distinction between strict coreference and internal/presumed/putative coreference (Fine 2010 ; see below, section 4). See also Goodsell 2014 on the distinction between coreference *de jure* and 'assumed coreference'.

[2] When two singular terms are coreferential *de jure*, Fine says that they '*represent* (their referent) *as the same*'. In contrast, an explicit identity statement such as 'Cicero is Tully' is said to represent the referent of the two singular terms as *being* the same. The two names are not coreferential *de jure* in the identity statement (see next footnote), so they do not 'represent their referent as the same'.

[3] In the case of identity statements such as 'Cicero is Tully', one *can* doubt that there is coreference between the two names, hence doubt the *truth* of the statement, even though one fully understands it.

> Suppose that you say "Cicero is an orator" and later say "Cicero was honest," intending to make the very same use of the name "Cicero." Then anyone who raises the question of whether the reference was the same would thereby betray his lack of understanding of what you meant (Fine 2007 : 40).

Taylor claims that proper names are *essentially* devices of coreference. Their role is to build, and exploit, 'chains of explicit coreference', participation in which guarantees the sharing of subject matter with other participants. 'What it is to intend to use an expression as a name', he says, 'is to use that expression with the intention of either launching or continuing a chain of explicit coreference' (Taylor 2003 : 10). When the same name is used twice, coreference is linguistically guaranteed : 'Tokens of the same name are guaranteed to corefer, if they refer at all' (Taylor 2003 : 14-15).[4] The fact that different objects may bear (what sounds superficially like) the same name is, from this point of view, an accident, irrelevant to the design of language. Such homonymous names have to be treated as distinct names for the purposes of logico-linguistic analysis (Kripke 1980) or, better perhaps, as distinct *expressions* (i.e. syntactic types) made up of the same name (Fiengo and May 1998, 2006).[5] 'Names are vocabulary items', Fiengo and May write, 'and expressions are syntactic items that may contain names… A name, qua lexical item, may in principle occur in *m*-many syntactic expression-types, each of which may have *n*-many token occurrences' (Fiengo and May 2006 : 14-17). When the speaker who makes two successive utterances of 'Cicero' 'intend[s] to make the very same use of the name "Cicero"' (as Fine puts it), he produces two occurrences of the same expression ; but a homonymous name (say 'Cicero' as the name of a cat) would be a different expression, made up of the same name.

In addition to the test suggested by Fine, there is another way of testing for coreference *de jure*. Two terms α and β are coreferential *de jure* just in case they licence a pattern of inference which John Campbell (1987) famously dubbed 'trading on identity' (TI) :

> *Trading on identity (TI)*
> α is F
> β is G
> Therefore, something is both F and G

Trading on identity is licensed when the same name occurs in both premisses, as in (1) below, or when the singular term in the second premiss is anaphoric on the singular term in the first premiss, as in (2).

(1) Cicero is F
Cicero is G
Therefore, someone is both F and G

(2) Cicero$_i$ is F

---

[4] In addition to the norm that *distinct occurrences of the same name corefer*, Taylor puts forward a second norm governing proper names : *occurrences of distinct name types refer to distinct objects*. The second norm will not play any role in my discussion.

[5] Fiengo and May generalize to all 'expressions' what Taylor says about names : 'All tokens of a given expression (…) corefer, *as a matter of grammar*' (Fiengo and May 2006 : 18). Coreference *de jure*, for them, is a matter of type identity in all cases. As we shall see (section 2), this view leads them to treat an anaphor and its antecedent as two distinct realizations of the same expression (the same syntactic type).

he$_i$ is G
Therefore, someone is both F and G

In the absence of either recurrence or anaphora, however, TI is not licensed : an additional identity premiss is needed to reach the conclusion, as illustrated by (3).

(3) Cicero is F
Tully is G
Cicero = Tully
Therefore, someone is both F and G

2. Recurrence

Fiengo and May claim that coreference *de jure* is a matter of recurrence (type identity) in *all* cases, and not merely in the cases in which the same proper name occurs twice. In general,

> Occurrences of an expression type corefer if they refer at all ; all occurrences in a given discourse will be coindexed, and hence coreferential as a matter of representation. (Fiengo and May 1998 : 381)

When the same expression recurs, it carries the same semantic value ; that's the general principle. This explains why two occurrences of the same name (qua syntactic expression) corefer *de jure*. But Fiengo and May take the recurrence account to apply to anaphora as well. They treat an anaphoric expression and its antecedent as *two distinct realizations of the same expression* (the same abstract 'syntactic' type), despite the lack of morphophonemic identity between them (Fiengo and May 1996: 137). In their framework, just as two distinct 'syntactic expressions' may involve the same lexical item (say, the same homonymous name), what counts as the same expression from the *syntactic* point of view may sometimes involve distinct lexical items (a name and a pronoun).[6] Both types of case are illustrated by sentence (4) :

(4) After their first meeting, Aristotle$_1$ invited Jackie$_2$ to spend a week on his$_1$ yacht. She$_2$ was then readings the *Metaphysics*, where Aristotle$_3$ says that knowledge is a basic human need.

In this sentence the numerical indices track type identity at the underlying syntactic level. The two occurrences of 'Aristotle' in (4) are tokens of distinct expressions (one, *Aristotle$_1$*, referring to the shipping magnate Aristotle Onassis, the other, *Aristotle$_3$*, referring to the ancient philosopher), while the anaphoric pronoun 'she' in the second sentence counts as the same expression as its antecedent in the first sentence, in virtue of being syntactically coindexed. As a result, Trading upon identity is licensed in the latter case (*Jackie$_2$*/*she$_2$*) but not in the former (*Aristotle$_1$*/*Aristotle$_3$*). (4) justifies the inference to (5) :

(5) Onassis invited someone who was then reading the *Metaphysics.*

[6] 'Two NPs may be occurrences of the same syntactic expression even though one may contain the name 'John' and the other the pronoun 'he' ; binding theory proceeds on this assumption' (Fiengo and May 1998 : 383).

But the inference from (4) to (6) is not licensed, because the two occurrences of the name ‘Aristotle’ bear different indices and do not count as tokens of the same expression :

(6) Someone for whom knowledge is a basic need invited Jackie on his yacht.

The idea that an anaphoric expression is the same expression as its antecedent is reminiscent of an early stage of transformational grammar, where an anaphoric pronoun was taken to be a transformation of a copy of its antecedent. 'Martin sold his car' was analysed as derived from 'Martin sold Martin's car' by a pronoun transformation, a condition of which is the syntactic identity between the pronominalized term and its antecedent. Because different individuals may be called 'Martin', the syntactic identity was taken to include the identity of the referential index: in 'Martin$_1$ sold Martin$_1$'s car', there is syntactic identity, but in 'Martin$_1$ sold Martin$_2$'s car' there isn't, so pronominalization (leading to 'Martin sold his car') is only possible in the former case. Generative linguistics has given up the idea that pronouns result from a transformation of (a copy of) their antecedent, but Fiengo and May retain the idea of a syntactic identity, corresponding to coindexing, between the pronoun and its antecedent.

Fiengo’s and May’s use of ‘expression’ is somewhat counterintuitive because (in the case of anaphora) it abstracts from issues of morphophonemic identity. Moreover, special problems may be thought to arise in connection with examples involving epithets or anaphoric descriptions (rather than anaphoric pronouns) :

(7) a. I met John$_i$/my new neighbour$_i$ the other day
b. The bastard$_i$ did not greet me

Do we want to treat the name ‘John’ (or the description ‘my new neighbour’) and the anaphoric description ‘the bastard’ as (two tokens of) *the same expression type* ? Aren't they, rather, two distinct expressions, despite being co-indexed? Clearly they are. Yet the issue is more complex than meets the eye. According to Patel-Grosz (2012), epithets like ‘the bastard’ in (7) actually are (null) *pronouns* modified by a nominal appositive. The logical form of (7b) would be

(7b*) *pro$_i$,* the bastard, did not greet me.

If this is right, then we could treat the null pronoun *pro$_i$* as the same syntactic expression as its antecedent, in Fiengo’s and May's sense, while acknowledging that the appositive description 'the bastard' is *not* the same expression as e.g. the antecedent description 'my new neighbour'. (Indeed, the two descriptions carry distinct meanings.)

Be that as it may, we don’t have to treat an anaphoric pronoun as the same expression as its antecedent to acknowledge that *recurrence* is what ultimately grounds coreference *de jure*. What recurs, arguably, is not (or not necessarily) a *linguistic* representation but a *mental* representation. Let us admit that ‘in cases of anaphora (as when I say ‘I saw John, he was wearing a bowler hat’), we can have two expressions representing an object as the same without the expressions themselves being the same’ (Fine 2007 : 41). This does not prevent us from accepting Fiengo’s and May’s point that there *is* identity, at a suitably deep level of analysis. There is, one might say, identity at the *conceptual* level — at the level of thought or logical form. The expressions ‘John’ and ‘he’ are not the same, in the ordinary sense of ‘expression’, but they are associated with the same conceptual representation, and that is what co-indexing indicates. Because of co-indexing, anyone who understands the utterance has to *re-deploy* the singular concept associated with the name when processing the anaphoric pronoun. Likewise, in (7) the antecedent ‘John’ (or ‘my new neighbour’) and the anaphoric

description ‘the bastard’ are associated with *the same singular representation*. Anyone who understands the utterance has to re-deploy the singular concept associated with the antecedent when processing the anaphoric description.

Placing the relevant identity at the conceptual level does not mean that it is not syntactically encoded. Sentences partially encode thoughts, and it is plausible that *recurrence constraints on conceptual elements* are encoded in the syntax of natural language. What recurs, however, is primarily a conceptual element (whether or not there is identity of expression at the linguistic level). That justifies shifting focus to the mental level and considering the associated representations directly.

Another consideration supports the shift to the cognitive level. Arguably, there is nothing specifically *linguistic* about trading on identity, coordination, etc. Suppose I hold a glass in my hand while looking at it. The glass looks dirty, and it feels cold. When, on the basis of my perceptual experience, I judge that the glass is cold and dirty, I trade upon the identity of the seen glass and the touched glass (Campbell 1987). Similarly, when I keep track of the glass from *t* to *t’* and judge that it is moving, I trade upon the identity of the object I perceive at the various times throughout the attentional episode (Evans 1981). That is already, at the most basic level, the phenomenon we are trying to elucidate. In other words : when I say ‘the glass$_1$ looks dirty, and it$_1$ feels cold’, what I express in language is a thought whose constituents already bear the relevant relation of coreference *de jure* to each other. Likewise, when, on the basis of my perception of the glass, I form the intention to drink from it, there is coreference *de jure* between the referential elements in the perceptual judgment and the intention based on it (Kamp 1990). Coreference *de jure*, even though it manifests itself in language, is first and foremost a phenomenon at the level of thought.[7]

3. Mental files

According to many authors in the recent singular thought literature, the singular concepts through which we represent particulars (e.g. our concept of Cicero) are best construed as *mental files binding together the subject’s predications concerning a given object.* Predications bound to the same file, being about the same object as a matter of representational architecture, are *eo ipso* ‘coordinated’ and license Trading on Identity. If there are two predicates *F* and *G* in the subject’s file for a given individual (or in one of his files if he has several), the predicates are coordinated and the subject can infer that there is an *x* which is both *F* and *G*. But to get that result it is not sufficient for the subject to predicate *being F* and *being G* of one and the same object. If the subject predicates *F* and *G* of the same object by thinking of it through two distinct files (e.g. because she mistakenly thinks there are two distinct objects, Cicero and Tully, or because she takes the point of view of someone who does), then Trading on Identity is blocked. The subject cannot justifiably infer that there is an *x* which is both *F* and *G*.

I have just mentioned the case of a subject who mistakenly thinks there are two distinct objects while there is only one. That is a ‘Frege case’, illustrated by the Babylonians’ thoughts about Hesperus and Phosphorus. In such cases, it is possible for a rational subject to ascribe contradictory properties to the object since, from the subject’s point of view, there are two distinct objects and no contradiction is internally detectable. Frege accounted for such cases by distinguishing sense from reference. The sense is the way the reference is presented

[7] See Kamp 1990: 47. When it comes to thought, Fine talks of ‘corepresentation’ rather than ‘coreference’. On the importance of the coreference *de jure* in thought, see James (1890 : 459) and Millikan (1997).

— the 'mode of presentation'. In Frege cases, there is a single reference but two distinct modes of presentation.

That there are two distinct modes of presentation in Frege cases is definitive of modes of presentation. Nothing is a mode of presentation unless it obeys what Schiffer calls 'Frege's Constraint' :

> Necessarily, if *m* is a mode of presentation under which a minimally rational person *x* believes a thing *y* to be F, then it is not the case that *x* also believes *y* not to be F under *m*. In other words, if *x* believes *y* to be F and also believes *y* not to be F, then there are distinct modes of presentation *m* and *m'* such that *x* believes *y* to be F under *m* and disbelieves *y* to be F under *m'*. Let us call this *Frege's Constraint* ; it is a constraint which any candidate must satisfy if it is to qualify as a mode of presentation. (Schiffer 1978 : 180)

Mental files satisfy the constraint, so they qualify as modes of presentation. Let us imagine a Babylonian, Hammurabi, who assertively entertains the thought that *Hesperus is visible in the evening but Phosphorus is not* (it is visible in the morning). He refers to Venus twice, by means of the singular terms 'Hesperus' and 'Phosphorus' which are associated, for him, with two distinct mental files. The two contradictory predications 'visible in the evening' (predicated of Hesperus) and 'not visible in the evening' (predicated of Phosphorus) are bound to distinct mental files and give rise to no contradiction within any of the two files. The contradiction remains internally undetectable, so the subject's rationality is not impugned. But if the singular terms were, in the subject's mind, associated with the same mental file, the two predications would be coordinated and the contradiction would be immediately apparent. Trading on Identity would be licensed, and the subject would have to face the conclusion that some object is both visible in the evening and not visible in the evening (a contradiction). A rational subject would therefore be led to retract one of the two contradictory predications. It follows that Frege's Constraint is satisfied : if *x* believes *y* to be F and also believes *y* not to be F, then there are distinct mental files *m* and *m'* such that *x* believes *y* to be F under *m* and disbelieves *y* to be F under *m'*.

In this framework, coreference *de jure* at the language level is to be accounted for in terms of deployment of the same file in thought. The identity which grounds coreference *de jure* is not the identity of the expressions but the identity of the mental file associated with them.

Note that, because they play the mode of presentation role, mental files must satisfy a transparency constraint: the subject must know when the same mental file is deployed twice, and when two distinct mental files are deployed. Transparency is what motivates the sense/reference distinction in the first place. The subject may not realize that two terms (e.g. 'Hesperus' and 'Phosphorus') refer to the same object, or that they refer to distinct objects. So reference is not epistemically transparent. In contrast, sense *must be* transparent. If modes of presentation are not transparent, there is no reason to move from pure referential talk to mode of presentation talk in the explanation of rational behaviour (e.g. the subject's assenting to 'Hesperus is visible in the evening' but not to 'Phosphorus is visible in the evening'). Sense is the level at which the subject's rationality can be assessed, and this entails that senses are transparent to the thinker.

4. The factivity issue

We started with a characterization of *de jure* coreference as a relation of coreference that holds in virtue of meaning. Because the coreference relation holds in virtue of meaning,

whoever grasps the meaning knows that the relation obtains. Coreference is guaranteed by meaning, by what the subject grasps, so it cannot fail to obtain. This corresponds to Fine's notion of *strict coreference*. Strict coreference is semantically required coreference — coreference that is required in virtue of one's semantic knowledge. Mere coreference is not enough : it must be part of the subject's semantic knowledge that there is coreference. Since knowledge is factive, strict coreference also is factive : if there is strict coreference between two singular terms M and N, there is eo ipso coreference between them.

But there is a snag. Sometimes, two terms are coreferential *de jure* by the standard tests (the subject presupposes coreference and trades upon identity) while in fact there is no coreference. According to Lawlor (2010), the existence of such cases shows that Fine's notion of strict coreference is not the right notion to capture the phenomenon. The phenomenon we are after, she says, is *internal* coreference. It is, *for the subject*, a priori that the two terms corefer. Internal coreference is the idea that coreference is presupposed by the subject, but this does not entail actual coreference. Internal coreference is not factive, Lawlor argues. This suggests that we should get rid of the following claim, central to Fine's characterization of coreference *de jure* as strict coreference :

> *Factivity:*
> Coreference *de jure* (in Fine's framework: strict coreference) entails coreference.

There are two types of case in which factivity seems to fail. First, there are empty singular terms. They do not refer, yet they can stand in *internal* 'coreferential' relations. Thus the subject can point to an object he hallucinates, attempt to designate it by e.g. 'that dagger', and then attempt to *de jure* corefer to the same object by uttering an anaphoric pronoun 'it' or by uttering an anaphoric description 'the dagger'. In such a case the subject trades upon identity in the normal way and coordinates his predications (so there is 'coreference *de jure*', by the standard tests), yet the failure of reference prevents coreference relations from actually obtaining. (Coreference entails reference, so, by contraposition, non-reference entails non-coreference.) The same sort of thing happens when one uses fictional names in discourse : a fictional name and a pronoun anaphoric on it bear the same referential index (they are associated with the same mental file), so they are coreferential *de jure*, by the standard tests ; yet, because neither of the singular term refers, they cannot corefer.

The second type of case in which factivity fails is the case in which the subject is confused. Lawlor gives the following example :

> Wally says of Udo, 'He needs a haircut', and Zach, thinking to agree, but looking at another person, says, 'he sure does'. (Lawlor 2010 : 4)

Here, Zach presupposes that, in his dialogue with Wally, the two occurrences of the pronoun 'he' corefer, but the presupposition is false. There is internal coreference for Zach since he trades upon the identity of the object he is looking at and the object antecedently referred to by Wally, but there is no actual coreference.

Fine agrees that internal coreference (what corresponds to the subject's own point of view) is not factive. The subject may treat two expressions as *de jure* coreferential, and behave accordingly (trading upon identity, etc.), even though the terms do not actually corefer. However, rather than giving up his (factive) characterization of *de jure* coreference as 'strict coreference' (a notion which entails actual coreference), Fine maintains it and attempts to define (nonfactive) internal coreference in terms of it. In other words, Fine advocates a view according to which there are *two distinct notions* of coreference *de jure* rather than a single one. Strict coreference corresponds to a first notion, that of semantically required

coreference. That is the basic notion. Two terms are strictly coreferential just in case it is part of one's semantic knowledge that they corefer. Knowledge is factive, so strict coreference entails actual coreference. The other notion is that of internal or putative coreference, which corresponds to the subject's point of view. Fine analyses it in terms of the basic notion. There is internal or putative coreference between two terms just in case the subject *treats them as* strictly corereferential, i.e. takes it for granted that they are. In the problematic cases, the subject is mistaken. There is no actual coreference and, therefore, no strict coreference either (since strict coreference entails coreference).

On Fine's picture, we must distinguish putative (or internal) coreference from strict coreference just as we distinguish strict coreference from mere *de facto* coreference. The three notions are built up as follows :

• **Mere (*de facto*) coreference** : a *and* b *corefer*.
COREF (a, b)
• **Strict coreference** : *it is part of the subject's semantic knowledge that (or : it is a semantic requirement that)* a *and* b *corefer*.
□ COREF (a, b)
• **Putative coreference** : *it is taken to be the case that* a *and* b *strictly corefer.*
T □ COREF (a, b)

Strict coreference is factive, but putative coreference is not. Here is Fine's own gloss on the trio of notions :

> Since the notion of being a semantic requirement is factive, the relation of strict coreference is likewise factive; strict coreference will imply coreference. And it is for this reason that the relation cannot be taken to be the relation of internal coreference. However, the notion of a putative semantic requirement is not factive; it can be a putative semantic requirement that P even though P is not the case. Suppose we take two singular terms to be putatively coreferential if it is a putative semantic requirement that they corefer. Then the relation of putative coreference is likewise not factive; two terms can putatively corefer without coreferring. And so there is no obstacle - or, at least, not the same obstacle - to taking this relation to be the relation of internal coference in cases of confused reference. (Fine 2010 : 497)

For Fine, coreference *de jure* is strict coreference, but Trading on Identity is not a an infallible test. What TI shows is that the subject treats the two terms as coreferential *de jure* (strictly coreferential). That does not entail that they are. In cases of delusion and confusion, they are not.

5. The transparency issue

Meaning is supposed to be transparent, i.e. known to the language users (Dummett 1978: 131). Since *de jure* coreferential relations are an aspect of meaning, it seems that they *must be* transparent to the language user. It must not be possible for the language user to be mistaken as to whether or not *de jure* coreference relations hold. But that is exactly what Fine says is possible in the case of strict coreference : it is possible for the subject to take it that there is strict coreference while in fact there isn't. This suggests that strict coreference is the wrong foundation for an account of coreference *de jure*.

The relevant notion of transparency is what I call Full Transparency :

Full Transparency (for a relation R)
For any two terms M and N, the subject knows whether or not M and N stand in the R relation to each other.

Full Transparency fails for strict coreference : the subject does not always know whether or not M and N are strictly coreferential.[9] This is what makes Fine's analysis of coreference *de jure* as strict coreference objectionable.

Fine himself can respond that he has, in his framework, a relation that is fully transparent : that is the relation of putative (internal) coreference. However, that is not the basic relation in terms of which Fine analyses coreference *de jure*. The basic relation is strict coreference. Because it is an aspect of meaning, that relation ought to be fully transparent; but, on Fine's analysis, it is not.[10]

Lawlor suggests *substituting* internal coreference for Fine's strict coreference in the analysis of coreference *de jure*. Coreference *de jure*, for Lawlor, *is* internal coreference. That relation can hold between two terms M and N even though M and N are not actually coreferential. The subject knows whether or not there is coreference *de jure*, but she does not thereby know whether or not the two terms actually corefer. Internal coreference is not factive, in contrast to strict coreference.

But I think one should not let factivity go too hastily. We are not interested in what the language users *take to be the case* (a nonfactive notion), but in what they *know to be the case* in virtue of their linguistic understanding. As Fine rightly insists, what we are after is a *semantic fact*, corresponding to the relation of *de jure* coreference. We must distinguish that semantic fact (something objective) from the subjective state a person is in when she knows that fact — the CDJ state, as I call it :

CDJ state
A subject is in the CDJ state with respect to two expressions M and N just in case she is disposed to presuppose coreference between M and N and to trade upon the identity of their reference.

Internal coreference (in the sense of Lawlor and Fine) holds between two terms just in case the subject is in the CDJ state with respect to them. Cases of emptiness and confusion show that a speaker may be in the CDJ state with respect to M and N even though M and N are not

---

[9] By 'the subject' I mean a competent and appropriately situated language user (typically the speaker herself, or an interpreter who properly understands the utterance).

[10] The mental file analysis delivers full transparency because of two 'Fregean' claims it rests on. First, an utterance is not understood unless the interpreter is able to associate the right modes of presentation with the referential occurrences of expressions in the sentence. The right modes of presentation are mental files in the interpreter's mind, meeting the constraints imposed by the meaning of the sentence and the common ground. (Such files are coordinated with the speaker's own files, which are subject to the same constraints). Second, modes of presentation are fully transparent : the subject knows when the same mental file is deployed twice, and when two distinct mental files are deployed (see section 3). The two claims together entail that a competent and properly situated interpreter will associate mental files with all referential expressions in the sentence and will know, for any two such files deployed in interpreting the utterance, whether or not they are occurrences of the same file. That means that, for any two referential expressions in the sentence, the competent and properly situated interpreter will know whether or not they are coreferential *de jure*, that is, associated with the same file and therefore bound to corefer.

actually coreferential. *But that does not mean that the notion of coreference* de jure *we are after is not factive*. There is, I claim, something which the speaker in the CDJ state (and anyone who understands the discourse) *knows* about the relation in which the two terms stand to each other. More precisely: there is a relation R such that, when the speaker is in the CDJ state with respect to a pair of terms M and N, he or she (and anyone who understands the discourse) knows that M and N stand in relation R. That relation R I call the *base relation* for de jure coreference. Coreference *de jure* is defined in terms of it:

> *Coreference* de jure *(schematic definition)*
> Two terms M and N are coreferential *de jure* $=_{\text{def}}$ The speaker is in the CDJ state with respect to them and, as a result, anyone who properly understands the discourse knows that the base relation R obtains between M and N.

Since coreference *de jure* involves knowledge, it is factive: whenever two terms M and N are coreferential *de jure*, they ipso facto bear the relation R to each other.

What of the counterexamples to factivity? They were only counterexamples to a specific way of construing the base relation R. The subject in the CDJ state cannot be said to know that M and N are R-related, if we take R to be the plain coreference relation;[12] but there are other candidates for the R relation than actual coreference. By suitably weakening the R relation, we can rescue factivity and obtain a notion of coreference *de jure* which is *both* factive and fully transparent. Following this path will make it possible to eschew Fine's split of coreference *de jure* into two distinct notions, one which is factive (but not fully transparent) and the other one which is fully transparent (but not factive).

6. Weakening the base relation

Fine himself (in passing) has mentioned different manners of construing the base coreference relation appealed to in the characterization of strict coreference. He writes the following in a footnote :

> Corefence between the names N and M may be defined *existentially* as $\exists x$(Ref(N, x) & Ref(M, x)) or *universally* as $\forall x$(Ref(N, x) ≡ Ref(M, x)). The two definitions are equivalent given that N and M have unique referents, i.e. given $\exists!x$Ref(N, x) & $\exists!x$Ref(M, x). (…) However, *when empty names are in question, it may be important to adopt the universal rather than the existential form of definition, since it will then be possible to distinguish between different empty names in regard to whether they strictly corefer.* (Fine 2007 : 134-35, emphasis mine)

The universal definition of coreference corresponds to Perry's notion of *conditional coreference* or 'coco-reference'. Taylor's notion of explicit coreference, Pinillos' notion of coreference *de jure*, Fiengo and May's notion of grammatically determined coreference, Gibbard' notion of assumed coreference, Goodsell's notion of IK-coreference, etc. — all work in the same way. In each case, the relation holds provided the two singular terms corefer *if they refer at all*. That conditional relation holds between the singular terms even if they fail to refer (in which case the antecedent of the conditional is false). This suggests the following characterization of coreference *de jure* as entailing knowledge of *conditional* coreference :

[12] As we have seen, M and N can be *de jure* coreferential (by the standard tests) even though M and N do not corefer.

Coreference de jure *(conditional characterization)*
Two terms M and N are coreferential *de jure* just in case : the speaker is in the CDJ state with respect to them and, as a result, anyone who properly understands the discourse knows that M and N corefer *if they refer at all.*

If we take the base coreference relation to be conditional coreference, then we can accommodate the empty cases (hallucination, fiction etc.) without having to split coreference *de jure* as Fine does when he distinguishes strict coreference (which is factive but not fully transparent) and internal/putative coreference (which is fully transparent but not factive). Under the conditional characterization, coreference *de jure* is factive. If M and N are coreferential *de jure*, it is a semantic fact that they corefer if they refer at all. That is true even if the terms are empty (in which case the antecedent of the conditional is false), so there is no need to give up factivity to dispose of the alleged counterexamples based on emptiness. When two singular terms are coreferential *de jure*, a competent and properly situated language user knows that they corefer if they refer at all, and that is compatible with the terms' failing to refer.

But there is the other class of counterexamples – the cases of referential confusion, as in the Wally/Zach story, repeated here:

> Wally says of Udo, 'He needs a haircut', and Zach, thinking to agree, but looking at another person, says, 'he sure does'. (Lawlor 2010 : 4)

In such a case, it seems that *even conditional coreference fails*. One cannot say that the two singular terms (the two pronouns) corefer if they refer at all. Certainly, the pronoun in Wally's mouth does refer (it refers to Udo). But it is implausible that the second pronoun (in Zach's mouth) also refers to Udo. Clearly, Zach is confused : he purports to refer not to Udo, but to another person he sees, whom he wrongly takes to be the person Wally was referring to. We can say either that Zach refers to that other person he sees, or that he fails to refer because he is confusedly tracking two distinct objects at the same time (namely the person he sees and the person Wally initially referred to). Whichever option we pick, the conditional coreference requirement is falsified : it is not the case that

$\forall x$(Ref(Wally's 'he', x) ≡ Ref(Zach's 'he', x))

It is because of such cases that we seem compelled to give up factivity and make room for a nonfactive notion of internal coreference.

One might discard that type of example on the grounds that, because of his confusion, Zach does not count as a competent and properly situated interpreter ; he does *not* understand Wally's utterance. Moreover, the example involves a dialogue : the two singular terms that are supposed to be coreferential *de jure* belong to the utterances of two distinct persons. Such cases introduce considerable complications, as two different points of view (and the mental files of two different subjects) come into play. These objections are well-taken, but I will not dwell on them, for there are other example of confusion which have the same structure but do not involve a dialogue, nor any form of linguistic deficiency. I will mention one in section 7, and a few more in section 9.

Confronted with such cases, we may appeal the same strategy we used to rescue factivity in the face of the empty cases. In the empty cases, what makes it possible to retain factivity is a shift in the base coreference relation in terms of which coreference *de jure* is defined. Instead of taking the base coreference relation to be the relation which holds between two terms just in case *there is* an object to which they both refer ($\exists$-coreference), we take it to

be that which holds between two terms whenever *any* object to which one of the terms refers is also referred to by the other term (∀-coreference). That base coreference relation is universal, not existential. Coreference *de jure* based on that coreference relation is compatible with failure of reference, so it can do the work of internal coreference when the terms are both empty. The subject in the CDJ state with respect to two terms M and N which turn out to be empty is still correct in treating the two terms as coreferential in the base sense, that is, as such that *if* one term refers to a certain object then the other term does as well. The idea, then, is to appeal to the same strategy in dealing with the cases of confusion, by weakening the base relation once again.

7. Weak coreference *de jure*

So far we have two candidates for the base relation R between two singular terms M and N in a situation of coreference *de jure*:

• The ∃-coreference relation: $\exists x(\mathrm{Ref}(\mathrm{N}, x) \,\&\, \mathrm{Ref}(\mathrm{M}, x))$
• The ∀-coreference relation: $\forall x(\mathrm{Ref}(\mathrm{N}, x) \equiv \mathrm{Ref}(\mathrm{M}, x))$

If we choose to base coreference de jure on the ∃-coreference relation, we cannot maintain factivity for coreference *de jure* because of both empty cases and cases of confusion. If we choose to base coreference *de jure* on the ∀-coreference relation, we can account for empty cases without giving up factivity, but cases of confusion are still counterexamples. To deal with these cases, we have to shift to an *even weaker* base relation. At this point, it will help to consider all the possible coreference options for two singular terms M and N.

There are four main types of (non-)coreference relation between M and N in a given piece of discourse (Table 1). Cases of type 1 corresponds to the ∃-coreference relation. If we base coreference *de jure* on that relation, we cannot account for cases of type 2 to 4 consistently with factivity. The ∀-coreference relation does better since it covers both cases of type 1 (∃-coreference) *and* cases of type 2 (the empty cases), but the notion of coreference *de jure* based on that relation still cannot be factive because of the possibility of cases of referential divergence (type 3 or 4) and not merely of referential emptiness.

Type 1. The two terms refer to the same object
Type 2. The two terms fail to refer
Type 3. One term refers to something, the other term fails to refer
Type 4. One term refers to something, the other term refers to something else

Table 1 : Possible (non-)coreference relations between two singular terms M and N

To accommodate (some of) the cases involving referential divergence, we can weaken the base relation a little more, by *reinterpreting* the claim that two terms are coreferential *de jure* only if they are known to *corefer if they refer at all*. For that claim is ambiguous : The conditional coreference relation ('corefer if they refer') can be interpreted in two ways, only one of which corresponds to ∀-coreference. The two interpretations are :

• Conditional coreference (strong) : M and N corefer if *either* refers (= ∀-coreference).
• Conditional coreference (weak) : M and N corefer if *both* refer.

These two base relations give us two alternative notions of coreference *de jure*, a weak one ('weak CDJ') and a strong one ('strong CDJ') :[13]

> *Strong CDJ:*
> Two terms M and N are coreferential *de jure* just in case : the speaker is in the CDJ state with respect to them and, as a result, anyone who properly understands the discourse knows that *M and N corefer if either refers.*
>
> *Weak CDJ:*
> Two terms M and N are coreferential *de jure* just in case : the speaker is in the CDJ state with respect to them and, as a result, anyone who properly understands the discourse knows that *M and N corefer if both refer.*

According to the *strong* notion of coreference *de jure*, a subject in the CDJ state with respect to M and N knows that

$\forall x(\text{Ref}(\text{N}, x) \equiv \text{Ref}(\text{M}, x))$

In words : the subject knows that if one of the term refers, the other term refers to the same thing. This rules out all cases of referential divergence between M and N, whether of type 3 or of type 4. Now it seems that a subject can be in the CDJ state with respect to M and N even though, because of confusion, one of the two terms (but not the other) fails to refer. Such cases are of type 3. Thus Wally's use of the pronoun 'he' refers to Udo, while Zach's partly anaphoric use of the anaphoric pronoun 'he' is confused and (arguably) fails to refer. Since the subject is in the CDJ state, we want to say that there *is* coreference *de jure* between the two pronouns ; but this is only possible (consistently with factivity) if we give the *weak* interpretation of coreference *de jure*. Only weak CDJ allows for cases in which one of the term refers and the other one fails to refer (type 3). With respect to such cases, the analysis of coreference *de jure* as strong CDJ would violate the factivity constraint. Factivity can be restored, however, by moving to weak CDJ. Weak CDJ is based on a relation R weaker than the ∀-coreference relation.[14]

According to the weak characterization of coreference *de jure*, a subject in the CDJ state with respect to M and N knows that if *both* terms refer, then they refer to the same thing. That is our third candidate for the status of base relation:

• The ∀∀-coreference relation : $\forall x \forall y$ ((Ref(N, x) & Ref(M, y)) → x = y)

---

[13] In section 9, we shall see that there are not merely two distinct *notions* of coreference *de jure* (the weak one and the strong one), but two distinct phenomena — two distinct *forms* of coreference *de jure*.

[14] Weak CDJ corresponds to what philosophers generally mean by 'coreference *de jure*'. Thus Pinillos says that a fully competent subject knows, of two occurrences M and N that are *de jure* coreferential, that they 'refer to the same object if the first refers to some object and the second refers to some object' ; they ' know of the expression occurrences that that if both have referents, then they refer to the same thing' (Pinillos 2011: 304). Goodsell's 2014 account of *de jure* coreference also equates it with weak CDJ: see her definition of 'IK-coreference' (Goodsell 2014 : 309). Drapeau-Contim (2016) is an exception : coreference *de jure*, for him, is strong CDJ, based on the ∀-coreference relation (see his principle of 'referential equivalence').

The ∀∀-coreference relation covers not only cases of emptiness but also cases of confusion (type 3), since in such cases the antecedent of the conditional is false : it is not the case that the two terms refer (since one of them fails to refer). Factivity is rescued, because the conditional is true, in such circumstances. If M and N are *de jure* coreferential (in the weak sense), then, whether or not the subject is deluded or confused, it is a fact that M and N corefer if they both refer.

Note that even weak CDJ rules out cases of type 4, i.e. cases in which M and N refer to two different things. Referential divergence is only allowed if one of the two terms fails to refer (while the other succeeds in referring to some object). When one of the terms fails to refer, the antecedent of the conditional ('if both refer') is falsified, so the conditional itself ('corefer if they both refer') remains true. This is consistent with factivity. But in a case of type 4, both M and N refer (so the antecedent is true) yet they do not corefer -- they refer to two distinct things. How can we dispose of such counterexamples? Can we weaken the base relation even more to accommodate cases of type 4?

Fortunately, we do not have to (or so I claim). Type 4 is not a problem, because there are no examples in which the subject is in the CDJ state with respect to two terms M and N yet M and N refer to distinct things. Cases of confusion fall either under type 2 or under type 3, but never under type 4. The reason is plain: *If the subject is confused, one at least of the two singular terms must fail to refer.* If I am right, there is no need to worry about type 4.

As I pointed out, the Wally/Zach case which may be thought to illustrate type 4 is controversial because it involves a dialogue. But consider a simpler, nonlinguistic example putatively of type 4. Imagine a subject who continuously tracks an object as it moves. At some point, unbeknown to the subject, a substitution occurs. The initial object, A, is replaced by a different object, B. Before the substitution, the subject's thought is uncontroversially about A. After the substitution, one might think that the subject's thought is about B (the new object). If it is, then the example falls under type 4. Presupposing the identity of the object B he is now tracking ('that is F') and the object A he was tracking before the substitution ('that was G a moment ago'), the subject trades upon the identity and infers : 'something which was G a moment ago is now F'. If this description of the case is correct, the subject is in the CDJ state with respect to the two demonstratives even though they refer to different things. Thus understood, the example falls under type 4. But this description neglects the fact that, after the substitution (but not before), the subject's thought is confused. Before the substitution, the subject's thought is about A. After the substitution, it is (partially) about B, *but it continues to be (partially) about A*, because it is presupposed that a single object is being tracked throughout the attentional episode. The presupposition is false and, as a result, the subject's attempted demonstrative reference to 'the' object arguably fails. This is a case of type 3 — the sort of case which the weak notion of coreference *de jure* allows.

The same considerations apply to the Wally/Zach case. Wally uncontroversially refers to Udo, but Zach's pronoun is linked deictically to the person he sees and anaphorically to Wally's pronoun, since Zach wrongly assumes that the person he sees is the person Wally was referring to. To the extent that Zach's pronoun targets two different individuals, it fails to refer simpliciter, contrary to Wally's pronoun. This reasoning works for all cases of confusion allegedly falling under type 4.

I conclude that, because confusion generates reference failure, it yields cases of type 2 or 3, but never of type 4. As a result, confusion cases do not threaten the factivity of coreference *de jure* understood as weak CDJ.

8. The transitivity issue

Accounting for coreference *de jure* in terms of identity of file entails that coreference *de jure* is a transitive relation (since identity is). But it is controversial whether coreference *de jure* is actually transitive, and there is an ongoing debate over precisely that issue (Pinillos 2011, Recanati 2012, Goodsell 2014, Drapeau Contim 2016).

Soames (1994: 253/2009 : 114) was the first to provide examples in which transitivity seems to fail :

(8) Mary told *John* that *he* wasn't John
(9) *John* fooled Mary into thinking that *he* was not John.

Transitivity seems to fail, for the following reason. We have seen that an anaphoric pronoun is coreferential *de jure* with its antecedent, so the first occurrence of 'John' in (9) is coreferential *de jure* with the pronoun 'he'. We have seen also that, in the normal course of events, two occurrences of the same proper name are coreferential *de jure*, so the first and the second occurrence of 'John' in (9) should be coreferential *de jure*. But the pronoun 'he' and the *second* occurrence of 'John' do not seem to be coreferential *de jure*. Frege's Constraint requires the existence of two distinct modes of presentation (distinct files) *m* and *m'* associated with the pronoun and the second occurrence of the name. Mary is said to have been fooled into thinking that John was not John. If Mary is rational, she must have thought of John under two distinct modes of presentation. But if there are two distinct modes of presentation, as in Frege cases, then the singular terms associated with these modes of presentation *cannot* be coreferential *de jure*. They can only be coreferential *de facto* (so the argument goes).

According to Soames, we should allow 'the term occurrences in the complement of (9) [to] be coordinated with the subject of 'fooled' without being coordinated with each other' (Soames 2010 : 474n). Since coordination is the same thing as coreference *de jure*, Soames' proposal amounts to the claim that A can be coreferential *de jure* with B and B with C, *without* A's being coreferential *de jure* with C. But if we accept that coreference *de jure* is not transitive, then, Pinillos (2011) points out, we can no longer claim that we can account for it by associating mental files with singular terms. We can no longer say that what accounts for trading on identity is the fact that *the same file* is deployed twice. If coreference *de jure* was a matter of identity (of files, of senses, or of whatever) it would be transitive – but it is not. That is an argument against *all* the 'third object' views, which account for the phenomenon of coreference *de jure* by positing a single entity associated with the two singular terms.

The argument is not compelling, however. In the Soames example (and several others constructed by Pinillos on the same pattern), the failure of transitivity is merely apparent. The appearance is due to a shift in point of view. From the speaker's point of view (or more generally, from the point of view of the speech participants), the pronoun and the second occurrence of the name (that which does not serve as antecedent to the pronoun) *are* coreferential *de jure*. The same mental file for John is deployed in association with the first occurrence of the name, the pronoun, *and* the second occurrence of the name. The speaker and the understanding hearer know that it is *John* who fooled Mary into thinking that *he* was not *himself*. The appearance that there is only coreference *de facto* between the pronoun and the second occurrence of the name comes from the illegitimate intrusion of another point of view, that of Mary, the person to whom an attitude is ascribed in (9). Being rational, Mary must think of John under two distinct modes of presentation (via distinct mental files) in order to believe of him that he is not John. So the pronoun and the second occurrence of the name correspond, in Mary's thought, to two distinct ways of thinking of John. But this pertains to

Mary's thought, not to the utterance. The mental files directly relevant to the interpretation of the utterance are those which *the speaker and her addressee* (and anyone who understands the utterance) associate with the singular terms. Mary's mental files are relevant only indirectly, because the utterance happens to report Mary's thoughts. (We shall see later that there are cases of 'oblique' reference in which the mental files of the ascribee are directly relevant to the interpretation of the singular terms ; but that is not the case in this example.)

In *Direct Reference* (Recanati 1993), I drew a distinction between two types of mode of presentation at work in singular attitude ascriptions. In a singular attitude ascription, a thought is ascribed to someone, about a particular object. The speaker refers to the object in reporting a thought about it. The way the speaker (and his addressee) think of the object is the 'exercised mode of presentation'. The exercised mode of presentation is the way the reference is presented in the discourse — the files which the speaker and the hearer deploy in mentally relating to the object the discourse is about. A certain way of thinking of the object is typically also *ascribed*, in a context-dependent and implicit manner, to the person whose thought is reported. That is the 'ascribed mode of presentation'. Now the fact that, *in the ascribee's thought,* there are two distinct modes of presentation of the object (e.g. John, in Soame's example) does not establish that *the speaker* also deploys/exercises distinct modes of presentation in referring to that object. The distinctness of the modes of presentation in the ascribed thought (Mary's) is therefore compatible with the uniqueness of the mode of presentation associated with the two name-occurrences and the pronoun anaphoric on one of them. Soames' examples, therefore, are compatible with the mental file account of coreference *de jure*, appearances notwithstanding. The files which matter when it comes to appraising whether or not two singular terms are associated with the same file are the files which the *speech protagonists* (not the characters whose thoughts are reported) deploy in referring to the objects the reported thoughts are about. They are the 'exercised modes of presentation'.

Sometimes, however, the mental files of some person distinct from the speaker, possibly the ascribee, are directly relevant to the interpretation of the singular terms and cannot be ignored by the theorist. Sometimes the speech protagonists *themselves* represent the object vicariously, via some file borrowed from some other person. This is a form of cognitive 'deference'. Think of identity statements like 'Hesperus is Phosphorus'. The speaker who says that knows that Hesperus and Phosphorus are a single planet, but what she says is not the same thing as what she would say if she used the reflexive : 'Hesperus is itself' (Safir 1998 : 141). Even though the speaker knows the identity and thinks of Hesperus/Phosphorus as a single planet (Venus), the files which the speaker deploys in her thought and speech are not two deployments of her inclusive VENUS file, but deployment of two distinct files, the HESPERUS file and the PHOSPHORUS file, which *correspond to the files in the mind of the subject who does not know the identity* (the hearer, perhaps). As Laura Schroeter insightfully writes, an identity statement 'is best understood as responding to a doubt about the identity' of the individuals who are said to be the same (Schroeter 2007 : 614n). In the Hesperus/Phosophorus case, the two singular terms are associated with the files through which the *unenlightened* thinks of Venus qua morning star and qua evening star. In such cases I say that there is *oblique reference*: the speaker refers to an object by deploying files 'indexed' to other people (by taking their point of view). The exercised mode of presentation, in such a case, is not a 'regular file' of the speaker's but a vicarious file

'indexed' to some other subject whose point of view the speaker temporarily espouses (Recanati 2012 : chapters 14 and 15).[15]

Oblique reference may target a third party rather than the hearer. The possibility of deploying files indexed to third parties in referring is illustrated by my old 'your sister' example (Recanati 1987 : 63). The speaker ironically says *'your sister' is coming over* and refers, by the description 'your sister' in quotes, to the person whom a third party takes to be the addressee's sister (but whom both the speaker and his addressee know not to be the addressee's sister). The file that is deployed in this case is a file about that person, containing the mistaken bit of information (that she is the addressee's sister). That file is indexed to the person the speaker is ironically mocking (the third party). It is not a 'regular file' in the mind of the speaker or the hearer, but a vicarious file used for essentially meta-representational purposes (to represent how other people represent things in the common environment).

In the 'Hesperus is Phosphorus' case (and in identity statements more generally), the enlightened speaker deploys files indexed to the unenlightened addressee, and refers obliquely (to Venus) through them. In a variant of that example, due to Pinillos, the target of oblique reference, i.e. the person to whom the vicarious files are indexed, is not the addressee but the speaker herself at an earlier time when she and her peers were not yet enlightened.

(10) We were debating whether to investigate both *Hesperus* and *Phosphorus* ; but when we got evidence of their true identity, we immediately sent probes *there*.

In the first clause the speaker espouses the point of view of the unenlightened (including herself before learning the identity), and she refers to Venus via two distinct mental files rather than via a single inclusive file corresponding to her current point of view. The inclusive file is associated with the demonstrative adverb 'there' at the end of the second clause. Again, the second clause expresses the subject's current point of view while the first clause is phrased from the point of view of the speaker and her peers before they learnt the identity. Since the first clause is a report ascribing certain speech acts to the speaker and her peers ('we were debating whether...'), the target of oblique reference in this case is the ascribee : the *exercised mode of presentation* associated with the two singular terms 'Hesperus' and 'Phosphorus' are mental files *indexed to the ascribee*, viz. the speaker and her peers at the time of the deliberation which is reported.[16]

To sum up, Pinillos's example involves three coreferential files : two indexed files (the HESPERUS file and the PHOSPHORUS file, indexed to the ascribee) and a regular file (the inclusive VENUS file, corresponding to the speaker's current point of view). They all have the status of 'exercised mode of presentation' because the speaker deploys each of them in referring, so they cannot be discarded as irrelevant, as the ascribed modes of presentation were in discussing Soames' example.

Because they are associated with distinct files, it is doubtful that the terms 'Hesperus' and 'Phosphorus' in Pinillos' example are coreferential *de jure*. Indeed, the usual test shows that they are coreferential only *de facto*: someone who disbelieves the identity could still understand the first clause. However, someone who disbelieves the identity would have trouble with the third term, 'there', which can only refer if the identity 'Hesperus = Phosphorus' is true. That term is associated with the speaker's inclusive file for Venus. As I

[15] Technically, there is no 'ascribed mode of presentation' in such an example, since it is not an attitude ascription. The same consideration applies to the next example of oblique reference (the 'sister' example). See *Mental Files*, pp. 201-202.

[16] In this special case, the indexed files play the role of *both* exercised and ascribed modes of presentation.

pointed out, the first clause is phrased from the point of view of the unenlightened (before learning the identity), while the second clause reflects the point of view of the enlightened, after discovering the identity and opening an inclusive file. The problem is that, even though the terms 'Hesperus' and 'Phosphorus' are only coreferential *de facto*, *each of them is coreferential* de jure *with the inclusive term 'there' in the second clause* (the term associated with the inclusive file) : the speech protagonists know that either 'there' fails to refer to a unique location (if the identity Hesperus = Phosphorus is not true), or (if the identity is true) it refers to the location of the single planet which Hesperus and Phosphorus turn out to be. That piece of knowledge corresponds to weak CDJ : for each of the two terms 'Hesperus' and 'Phosphorus', the subject knows that *that term corefers with the the inclusive term 'there' if they both refer*. In other words : 'Hesperus' is in the (weak) CDJ relation to 'there', 'Phosphorus' is in the (weak) CDJ relation to 'there', yet 'Hesperus' and 'Phosphorus' do not stand in the weak CDJ relation to each other : they are not coreferential *de jure*, but *de facto*. This, Pinillos argues, shows that coreference *de jure* is not transitive. The proper representation of example (10), with coindexing, is

(10) We were debating whether to investigate both Hesperus$_{1}$ and Phosphorus$_{2}$; but when we got evidence of their true identity, we immediately sent probes there$_{1,2}$.

Pinillos gives another, particularly interesting variant of the example :

(11) Hesperus$_{1}$ is Phosphorus$_{2}$ after all, so Hesperus-slash-Phosphorus$_{1,2}$ must be a very rich planet.

In the first clause, the speaker espouses the point of view of the unenlightened and deploys the HESPERUS and PHOSPHORUS files. In the second clause the speaker shifts to her current, enlightened point of view, and deploys the inclusive file. So far, this is like the previous example, but what is interesting about this variant is the term which is associated with the inclusive file in the second clause : 'Hesperus/Phosphorus'. That is a complex term, sometimes referred to as a 'slash-term'. Slash-terms are composed of two terms (the 'basic terms', here 'Hesperus' and 'Phosphorus') plus the slash operator.[18] What the slash operator does is create a new term which refers to the same thing as each of the basic terms if they corefer, and to nothing otherwise. The slash-term presupposes that the basic terms corefer, and itself refers only if the basic terms do corefer. It follows that *weak CDJ to each of the basic terms is a built-in feature of slash-terms*. Anyone who masters the slash-term knows a priori that either it corefers with each of the basic terms or it fails to refer. That is sufficient to support weak CDJ between the slash-term and each of the basic terms. Since the basic terms are only coreferential *de facto*, transitivity fails for weak CDJ : the two basic terms A and B are each coreferential *de jure* with the slash-term A/B, yet A and B themselves are not coreferential *de jure*.

In *Mental Files*, I acknowledged the failure of transitivity, and I weakened the theory accordingly. The Pinillos examples reveal that, for two terms to stand in the weak CDJ relation, it is not necessary for them to be associated with the same file. Two terms will also stand in the weak CDJ relation if one is associated with an initial file, and the other with an inclusive file resulting from the *fusion* of that initial file with another initial file presupposed to be coreferential with it. Being associated with the same file is therefore a sufficient condition for two terms to be coreferential *de jure*, but it is not a necessary condition.

[18] The label 'basic term' comes from Drapeau Contim (2016).

I still accept what I said in *Mental Files*, but I think we should pay more attention to the distinction between weak CDJ and strong CDJ, construed now as two *forms* of coreference *de jure* (rather than two competing *conceptions* of what coreference *de jure* is). Weak CDJ, I maintain, is not transitive, and it cannot be equated to the relation of being associated with the same file (that would make it transitive, as Pinillos points out). Being associated with the same file is a sufficient condition for weak CDJ between two terms, but not a necessary condition. But I would like to suggest that strong CDJ *is* transitive, and *can* be equated to the relation of being associated with the same file. So we don't really need to weaken the theory (as I did in *Mental Files*), we only need to carefully distinguish between the two forms of coreference *de jure*, weak and strong.

9. Strong coreference *de jure*

We moved from strong CDJ (based on ∀-coreference) to weak CDJ (based on ∀∀-coreference) because of cases of confusion. The first case of confusion we encountered was the Wally-Zach case due to Lawlor. The subject is in the CDJ state with respect to the two utterances of 'he', yet the first occurrence refers to Udo while the second one, being based on confusion, fails to refer (it simultaneously tracks Udo and the person Zach is seeing). Wally utters the first 'he', and refers to Udo. Zach utters the second 'he', referring to the man he sees, who he wrongly takes to be the person Wally was referring to. Zach's utterance of 'he' is both deictic and anaphoric, it seems. The direct referential link to the person seen does not prevent Zach's pronoun from being anaphorically linked to Wally's : that is made possible by Zach's presupposition that Wally's referent is the person Zach sees. This was presented as a 'type 3 case', i.e. a case in which one term (Wally's 'he') refers to one thing (Udo), while the other term (Zach's 'he') fails to refer due to confusion (section 6). But that example raises difficulties of its own, due to the fact that it involves a dialogue : the two occurrences of the pronoun 'he' that are supposed to be coreferential *de jure* are not uttered by the same person. Wally, indeed, refers to Udo, but Zach takes Wally's pronoun to refer to the man he sees, and it is with the pronoun *thus construed* that Zach intends to corefer *de jure*. If we consider only Zach's point of view, the mental file *he* associates with Wally's utterance of 'he' is the same mental file he associates with his own utterance of 'he', namely the confused file resulting from his mistaken identity presupposition.[19] It follows that, *from Zach's point of view*, the two pronouns corefer *de jure* in the strong sense : they corefer if either refers.

In section 6 I said there were other 'type 3' examples, not involving a dialogue but exhibiting the same structure as the Wally-Zach example. In section 7, I mentioned one that does not involve language at all. From time t_1 to t_3, the subject tracks an object as it moves, but fails to detect a substitution occurring at t_2, in the midst of the tracking episode. Before t_2, the subject's demonstrative file refers to A, the object tracked between t_1 and t_2. The object tracked between t_2 and t_3 is B, not A, but the file rests on the presupposition that one and the same object is being tracked throughout the attentional episode. Borrowing an idea from Hartry Field (Field 1973 ; see also Devitt 2015) we can say that the file after t_2 partially refers to B and partially refers to A. A deployment of the file before t_2 therefore refers to A, while a deployment of the 'same' file after t_2 refers to both A and B and so fails to refer simpliciter, the world failing to cooperate. In such a case, as in the more spectacular cases of mistaken fusion of files (e.g. Marco Polo's confusion regarding Madagascar), the weak CDJ relation holds between the initial file (before t_2) and the more inclusive file deployed after t_2.[20] I

[19] I am indebted to Philippe Lusson here.

[20] When I speak of the weak CDJ relation as holding between two files (rather than between terms), I mean that the subject who deploys the files knows that they corefer if both refer.

describe the file after t_2 as 'more inclusive' because it stores information from both A and B, and it rests on more information channels (the subject after t_2 remembers how 'the object' was before while perceiving how 'it' now is).

Because, in that example, the inclusive file presupposes the identity of the object tracked between t_1 and t_2 and the object tracked between t_2 and t_3, the structure of the example is similar to that of Strawson's cases of 'merging'. According to Strawson (1974), when one learns an identity, one merges the files one initially had. The inclusive file resulting from merging the two initial files rests on a presupposition of identity (it is, in Drapeau-Contim's terms, 'identity-dependent'). If the identity fails to hold, the inclusive file (and the slash-term possibly associated with it) fails to refer, but that does not prevent each of the initial files from referring. Both the infelicitous tracking case and the mistaken merge case are of 'type 3', the type of case that justifies moving from strong CDJ to weak CDJ.

Another type of example with the same structure involves mistaken recognition. One remembers a certain object A, and upon encountering an object B wrongly recognizes it as A. The initial memory file is about A, while the post-recognition file mixes memory information about A and perceptual information about B, presupposing that A and B are the same. This is an instance of what I call 'incremental conversion' (Recanati 2012, 2013, 2017): a file grows new information links as time passes, and its continued existence rests on the presupposition that all the information derives from the same object. If the identity presupposition is false, that is, in this case, if $A \neq B$, the initial memory file and the more inclusive recognitional file (hosting information derived from both memory and current perception) diverge in their referential status : the initial memory file refers to A, while its post-recognitional continuation fails to refer, due to confusion. Even though they are, in a dynamic sense, the 'same file', they are better seen as two distinct *file-stages* with divergent referential fates : the memory file before t_2 refers to A, while the inclusive file after t_2 fails to refer. Whenever that structure is instantiated, strong CDJ fails : it is not true that the two files 'refer if either refers'. One refers, but the other doesn't. Rather, the two files stand in the weak CDJ relation : although the identity presupposition is mistaken, the subject in the CDJ state still knows that the two files corefer if they both refer.

It is worth noting that, in all these nonlinguistic examples of confusion, the files in the weak CDJ relation are *deployed at different times*. Merging is the process through which an inclusive file substitutes for two initial files, which Strawson describes as 'withdrawn' (Strawson 1974 : 56). Likewise, the recognitional file supersedes the memory file, and the demonstrative file after t_2 supersedes the earlier file-stage (the demonstrative file before t_2). All these cases are diachronic, and it is a general fact that files deployed at different times can only support the weak CDJ relation. Because confusion can always arise as information is collected across time, file stages deployed at different times — even deployments of the same dynamic file, as in the infelicitous tracking example — cannot be coreferential *de jure* in the strong sense. That is most obvious when the file undergoes fusion or fission, but that holds also for the simpler types of example, without fusion or fission.

What about files deployed at the same time — synchronous files ? Unless distinct points of view are brought together, as when indexed files come into the picture, the situation is rather neat. Two synchronous deployments are either deployment of the same file, or deployment of distinct files. Two synchronous deployments of the same file stand in the *strong* CDJ relation to each other — they are known to corefer if either refers —, while deployments of distinct files can only be coreferential *de facto* (it may be that one file refers to A while the other refers to B, as in type-4 cases). With diachronic deployments of files the situation is a lot messier because of the possibility of confusion (type 3), due to failure of the world to cooperate. Of course, confusion is *always* possible, but not all confusion has to be of type 3. Confusion of type 2 is compatible with strong CDJ, and that is the sort of case we

encounter with synchronous files. If the subject is confused, the file she deploys fails to refer, *and fails to refer on all of its (synchronous) deployments*. No referential divergence is therefore generated when the file is deployed twice. This suggests that the shift from strong CDJ to weak CDJ can be avoided in the case of synchronous files.

What I said of the Wally-Zach *interpersonal* case can now be extended to the *diachronic* cases. The Wally-Zach case involves two points of view : that of Zach (the person in the CDJ state) and that of Wally. I said that if we focus on Zach's point of view we find that he associates *the same file* with the two pronouns respectively uttered by Wally and by himself. Weak CDJ characterizes the case in which different points of view are mixed, but if we fix the point of view strong CDJ is restored. The same considerations apply to the diachronic cases. In the diachronic cases, the two file-stages that are said to stand in the weak CDJ relation correspond to different temporal points of view. If we *fix* the temporal point of view and focus on e.g. the subject after t_2 (in the infelicitous tracking example), strong CDJ is restored. Even if the subject (after t_2) thinks of the object he was perceiving before t_2, he will think of it under the confused file resting on the mistaken identity presupposition : so the two deployments of the inclusive file (in thinking of the object as it was before t_2, and in thinking of the object as it is now) corefer if either refers. (Since we assume that the identity is mistaken, both deployments fail to refer ; this is a case of type 2, not type 3).

To be sure, the subject may attempt to refer *specifically* to the object perceived before t_2 (rather than the object she is currently perceiving), if she entertains the suspicion that a substitution may have occurred. In that case, however, she will *split* the inclusive file and refer through to two distinct 'daughter files'. I described such a case of fission in my paper 'Cognitive Dynamics' :

> At t_1, I see a certain object and open a demonstrative file DEM_1 about it : 'that thing'. At t_2, the object I have been in contact with since t_1 disintegrates, but the demonstrative file persists because, as a result of taking a certain drug, I hallucinate the continued presence of the object. (…) At t_3 a doubt occurs to me and I wonder whether the object I remember seeing at the beginning of the episode (t_1) is really the same as the object that I (mistakenly) take myself to seeing at t_3. Rational doubts about identity necessarily involve two distinct mental files, and here the two files result from splitting DEM_1, which is replaced by a memory demonstrative (referring to the object initially seen) and a perceptual demonstrative (purporting to refer to the object currently seen). (Recanati 2017 : 188)

In this example, there are, in diachrony, four different file-stages to consider. The demonstrative file opened at t_1 and maintained until t_3 (despite the disappearance of the object at t_2), is deployed both before and after t_2. Before t_2 it refers to A. Between t_2 and t_3 it fails to refer (for two reasons : the subject is hallucinating, and she wrongly presupposes that the object she wrongly takes herself to see is the same she has been perceiving all along). At t_3 fission occurs and two new files come into existence : a memory file about the object initially seen, and a demonstrative file about the object the subject hallucinates. The memory file refers to A (the object the subject remembers seeing) and the new demonstrative file fails to refer (because the subject is hallucinating). Let us call the four file-stages α, β, γ and δ. File α is deployed before t_2, and file β in the interval between t_2 and t_3. Files γ and δ are deployed after t_3 (the time of the split). Now, what are the coreference relations between the four file-stages ? The initial file α stands in the weak CDJ relation to its successor β, since β embodies a fallible presupposition of identity. File β is, with respect to α, an 'inclusive', identity-dependent file, susceptible to type 3 cases of mistaken identity. So files α and β cannot stand in the strong CDJ relation ; they can only instantiate weak CDJ. Files γ and δ result from

splitting β, and they are not coreferential *de jure* at all : as I have described the case, the subject suspects that they might not corefer, which is why the inclusive file β was split in the first place. However, file γ and file δ each stand in the relation of weak CDJ to file β : if the identity presupposed by β is mistaken, β fails to refer but that does not prevent γ (and, for all the subject knows, δ) from referring.

If we fix the temporal point of view, instead of looking at the coreference relations of files deployed at different times, strong CDJ is immediately restored. Between t_2 and t_3 the confused subject will deploy β *both* to think about the (hallucinated) object she takes herself to be seeing and to think about A, the object initially seen. The reason is that, after t_2, *α is no longer available to think about A* – it has been superseded by β. So instead of two deployments of distinct files standing in the weak CDJ relation, we have two deployments of the same file β. These deployments, and the singular terms associated with them, are coreferential *de jure* in the strong sense : they corefer if either refers. (Actually, they don't refer.)

These examples, as described, do not involve language, and the CDJ relations hold directly beween files rather than between singular terms. But we can let the subject speak ! Between t_2 and t_3 the subject might say :

(12) That object$_\beta$ was F but it$_\beta$ is now G.

After t_3, when she starts doubting, the subject might say :

(13) I wonder whether (that object)$_\gamma$, which was F, is (that object)$_\delta$, which is G, or whether a substitution occurred unbeknown to me.

In (12), 'that object' and 'it' are associated with the same file, namely the confused file β. So they are coreferential *de jure* in the strong sense. No diachrony is involved, even though the first clause, in the past tense, talks about the situation between t_1 and t_2, and the second clause talks about the current situation (after t_2). No diachrony is involved because the two deployments of β (in association with the demonstrative and with the pronoun) are synchronous deployments, or 'co-deployments'. In (13), however, the subject refers to the same putative object(s) by deploying two distinct files γ and δ. These files, and the terms they are associated with, are not coreferential *de jure* at all.

I conclude that synchronic deployments of the same file exhibit strong coreference *de jure*, while diachronic deployments, and the dynamic files (sequences of file-stages) they give rise to, only support weak CDJ.[22] Weak CDJ is intransitive, while strong CDJ is transitive.

[22] One of the editors of this volume, Rachel Goodman, ''wonder[s] if it is in the end true that only diachronic cases can give rise to weak CDJ. The case I have in mind is a synchronic cross-modal perceptual case. I touch a certain object while seeing it judging, at the same time, on the basis of touch, '*this* is rough', on the basis of sight, '*this* is red', and concluding 'this is red and rough'. Spelled out in the right way, this case might be both synchronic and involve weak CDJ.'' However, in several papers (Recanati 2013, 2017), I argued that *if* we construe the crossmodal reasoning as synchronic, *then* a single mental file is associated with the three occurrences of 'this', namely an inclusive file hosting both perceptual and tactile information. Because they are associated with the same mental file, the three occurrences are coreferential *de jure* in the strong sense : they corefer if either of them refers. It is only if the subject starts doubting whether the object touched is the object seen that he will split his inclusive file into two distinct mental files. If he does, however, the

Now, when we analyse an utterance or a thought, which notion of coreference *de jure* between constituents of the utterance or thought should we use ? Answer : the strong one, because the files associated with distinct constituents in an utterance are *codeployed in the thought which is the output of the interpretation process*. Either it is the same file that is codeployed, and the two deployments stand in the strong CDJ relation, or it is distinct files and the coreference is, at best, *de facto*.

The only exception to that principle is the case in which distinct points of view are simultaneously at play, in the interpretation of an utterance in which one or several terms are associated with indexed files. In that type of case, strong CDJ is still ruled out between the terms associated with distinct files, but the weaker form of coreference *de jure* enjoyed by diachronic deployments is now available in synchrony, via the mechanism of indexed files. That type of case, to which Pinillos has drawn our attention, is important because it reveals the intransitivity of the weak CDJ relation, but it is *not* a counterexample to the general principle I want to reassert (and leave as a take-home message to the reader) : two terms are coreferential *de jure* (in the strong sense) if and only if they are associated with the same file.[23]

References

Campbell, J. (1987) Functional Role and Truth Conditions: Is Sense Transparent ? *Proceedings of the Aristotelian Society,* Supplementary Volume LXI, 273–292.

Devitt, M. (2015) Should Proper Names Still Seem so Problematic ? In A. Bianchi (ed.) *On Reference*, pp. 108-43. Oxford : Oxford University Press.

Drapeau Vieira Contim, F. (2016) Mental Files and Non-Transitive De Jure Coreference. *Review of Philosophy and Psychology* 7 : 365-88.

Dummett, M. (1978) *Truth and Other Enigmas*. London : Duckworth

Evans, G. (1981) Understanding Demonstratives. In H. Parret and J. Bouveresse (eds.) *Meaning and Understanding*, pp. 280-303. Berlin : De Gruyter

Fauconnier, G. (1974) *La Coréférence: Syntaxe ou Sémantique?* Paris: Seuil.

Field, H. (1973). Theory Change and the Indeterminacy of Reference. *Journal of Philosophy* 70 : 462-481.

Fiengo, R. and May, R. (1996) Anaphora and Identity. In S. Lappin (ed.) *Handbook of Contemporary Semantic Theory*, pp. 117-44. Oxford: Blackwell.

Fiengo, R. and May, R. (1998) Names and Expressions. *Journal of Philosophy* 95 : 377-409.

Fiengo, R. and May, R. (2006) De Lingua *Belief*. Cambridge, Mass. : MIT Press/Bradford Books.

Fine, K. (2007) *Semantic Relationism*. Oxford : Blackwell.

Fine, K. (2010) Reply to Lawlor's 'Varieties of Coreference'. *Philosophy and Phenomenological Research* 81: 496-501.

Gibbard, A. (2012) *Meaning and Normativity*. Oxford : Oxford University Press.

Goodsell, T. (2014) Is De Jure Coreference Non-Transitive? *Philosophical Studies* 167: 291-312.

James, W. (1890) *Principles of Psychology*. New York : Holt.

---

subject will no longer be in a position to 'trade upon identity' and conclude 'this is red and rough'.

[23] My work on mental files, including this paper, has been supported by the French Agence Nationale de la Recherche under grant agreement n° ANR-10-LABX-0087 IEC and grant agreement n° ANR-10-IDEX-0001-02 PSL.

Kamp, H. (1990) Prolegomena to a Structural Theory of Belief and Other Attitudes. In C.A. Anderson and J. Owens (eds.) *Propositional Attitudes*, pp. 27-90. Stanford: CSLI.
Kripke, S. (1980) *Naming and Necessity*. Oxford : Blackwell.
Lawlor, K. (2001) *New Thoughts About Old Things*. New York : Garland.
Lawlor, K. (2010) Varieties of Coreference. *Philosophy and Phenomenological Research* 81 : 485-501.
Millikan, R. (1997) Images of Identity : In Search of Modes of Presentation. *Mind* 106 : 499-519.
Neale, S. (2005) Pragmatism and Binding. In Z. Szabo (ed.) *Semantics versus Pragmatics*, pp. 165-285. Oxford : Clarendon Press.
Patel-Grosz, P. (2012) *(Anti-)Locality at the Interfaces*. PhD dissertation, MIT.
Perry, J. (2012) *Reference and Reflexivity*, 2nd Edition. Stanford : CSLI.
Pinillos, A. (2011) Coreference and Meaning. *Philosophical Studies* 154: 301-24.
Recanati, F. (1987) Contextual Dependence and Definite Descriptions. *Proceedings of the Aristotelian Society* 87 : 57-73.
Recanati, F. (1993) *Direct Reference : From Language to Thought*. Oxford : Blackwell.
Recanati, F. (2012) *Mental Files*. Oxford : Oxford University Press.
Recanati, F. (2013) Perceptual Concepts : In Defence of the Indexical Model. *Synthese* 190 : 1841-55.
Recanati, F. (2017) Cognitive Dynamics : A New Look At an Old Problem. In M. de Ponte and K. Korta (eds.) *Reference and Representation in Thought and Language*, pp. 179-94. Oxford : Oxford University Press.
Safir, K. (1998) Abandoning Coreference. In Bermudez, J.L. (ed.) *Thought, Reference, and Experience*, pp. 124-163. Oxford : Oxford University Press.
Schiffer, S. (1978) The Basis of Reference. *Erkenntnis* 13 : 171-206.
Schroeter, L. (2007) The Illusion of Transparency. *Australasian Journal of Philosophy* 85 : 597-618.
Soames, S. (1994) Attitudes and Anaphora. *Philosophical Perspectives* 8 : 251-72.
Soames, S. (2009) *Philosophical Essays, vol. II : The Philosophical Significance of Language*. Princeton : Princeton University Press.
Soames, S. (2010) Coordination Problems. *Philosophy and Phenomenological Research* 81 : 464-74.
Strawson, P. (1974) *Subject and Predicate in Logic and Grammar*. London : Methuen.
Taylor, K. (2003) *Reference and the Rational Mind*. Stanford : CSLI.