Transparent Coreference

François Recanati

Transparent Coreference

François Recanati

Topoi (2019), https://doi.org/10.1007/s11245-019-09674-1

1. Referential content vs cognitive content

It is natural to consider that, to be a sign, and therefore to carry meaning, something has to *stand for* something else. A sign, the Port-Royal logic says, is an entity (for example a linguistic expression) that represents or stands for another entity. In contemporary philosophy of language, the relation of representing or standing for is called the reference relation. The entity the sign refers to itself is called its reference (or, less ambiguously, its referent).

Should we equate meaning and reference ? Although it is tempting to do so, there are well-known objections to that move. First, if meaning is reference, then an expression which fails to refer is bound to be meaningless ; but, to take a famous example, a complex expression like 'the present king of France' is meaningful even though, arguably, it fails to refer (since France is a Republic). In response to that objection, it is advisable to restrict the scope of the claim that meaning is reference (the 'basic equation', as I will henceforth call it) to *simple* expressions, e.g. 'king' or 'France'. The noun 'king' refers to the property of being a king, and the proper name 'France' refers to a particular country, namely France. To say what these expressions refer to is to say what they mean (or so it seems). To be sure, there are simple expressions, like the name 'Vulcan', which fail to refer yet seem meaningful. But it can be argued that such expressions only seem meaningful to those who are subject to the illusion that they refer (e.g. Le Verrier), or to those who take the perspective of someone who is subject to such an illusion (e.g. those who report on Le Verrier's beliefs). The first objection to the equation of meaning and reference is therefore not as convincing as many people — including one reviewer for this journal — seem to think.

A second objection to the basic equation is due to Frege and goes like this. If meaning is reference, then two coreferential expressions must carry the same meaning. Yet a rational subject can entertain contradictory attitudes towards two sentences that only differ by the substitution of one of two coreferential expressions for the other. Thus a rational subject may endorse 'Cicero was Roman' while rejecting 'Tully was Roman', i.e. they may believe the content of one sentence while disbelieving the content of the other. How is this possible if the content (= meaning) of the two sentences is the same ? And how could it not be the same, on the assumption that meaning is reference, if the two expressions 'Cicero' and 'Tully' corefer ?

Based on that objection, Frege introduced a distinction between two aspects, levels or dimensions of meaning (sense and reference), thereby rejecting the basic equation (Frege 1967). The additional dimension, sense, corresponds to the way the reference is presented. 'Cicero' and 'Tully' refer to the same individual, but carry distinct senses : the individual is presented differently in the two cases – as Tully, or as Cicero. So-called 'Frege cases' are cases like the Cicero/Tully example in which a rational subject takes conflicting attitudes towards what is in effect the same content, referentially individuated. What makes it possible for a rational subject to take these conflicting attitudes is the fact that the content in question is presented differently, through distinct senses. The modes of presentation associated with the names 'Cicero' and 'Tully' contribute to the utterance's *cognitive*

content, where cognitive content departs from referential content by obeying the following *constraint on cognitive content* (CCC) :

> (CCC) If a rational and linguistically competent subject believes a sentence S to be true yet believes a sentence S' to be false, then S and S' differ in cognitive content for that subject.

If, as in Frege cases, S' results from substituting an expression N for another, coreferential expression M occurring in S, then S and S' differ in cognitive content even though they share the same referential content. According to Frege, that difference in cognitive content can be traced to the difference in the modes of presentation of their common referent which are respectively carried by the two coreferential expressions.

An utterance's referential content corresponds to its *truth-conditions*, while its cognitive content corresponds to *the thought it expresses*. CCC follows from a basic *rationality constraint on thoughts* (RCT), to the effect that

> (RCT) A rational subject cannot (without ceasing to be rational) endorse and reject the same thought simultaneously.

That constraint entails Frege's cognitive criterion of difference for thoughts :

> Two thoughts T1 and T2 are distinct if it is possible for a rational subject to endorse T1 while rejecting T2.

This in turn entails CCC : if a rational and linguistically competent subject takes S to be true and S' to be false, they associate distinct thoughts with S and S' — they assign them distinct cognitive contents.

2. The transparency of cognitive content

What is characteristic of Frege cases is the fact that the two coreferential expressions at stake are not *known* to corefer. The subject who believes that Cicero was Roman but that Tully was not ascribes contradictory properties to what is in fact the same individual, but her rationality is not impugned because she does not know that Cicero is Tully. If she did, she would not accept 'Cicero was Roman' while rejecting 'Tully was Roman' : that would be irrational.

When our subject says 'Tully was not Roman', she does refer to Cicero (by means of the name 'Tully') but she is unaware that she is referring to Cicero, because she takes 'Tully' to refer to an individual distinct from Cicero. So, in a sense, she is referring to Cicero *without being aware of doing so*.[1] The fact that that is possible — that one can refer to an entity A without being aware of doing so — means that reference is not *transparent* to the subject. Which entity one is referring to depends upon the world, upon contingent facts of which the subject (who is rational, but not omniscient) may not be cognizant. As a result, *coreference* is not transparent to the subject either : it is possible for the subject to refer to the same

[1] Of course, as one reviewer emphasizes, there is *also* a sense in which she is aware of what she is doing.

individual twice (as in Frege cases) without knowing that the two acts of reference target the same individual. In such cases there may be an objective contradiction at the level of referential content : the subject may ascribe contradictory properties to one and the same individual. Internally, however, there may be no detectable contradiction, as reference is not transparent to the thinker. The subject only has transparent access to the mode of presentation – and here the modes of presentation are distinct. (If the modes of presentation were the same, there would be an internal inconsistency.)

The crucial point is that, while referential content is not transparent to the subject, cognitive content is. If *p* is the referential content carried by an utterance, then, as we have seen, the subject who accepts *p* may well, at the same time, reject *p* or accept *not-p*, provided the referential content *p* is accessed through distinct modes of presentation when accepting, and when rejecting. But if *p* is the cognitive content carried by an utterance, then, in virtue of RCT, the subject who accepts *p* cannot at the same time reject *p* or accept *not-p*. That would be irrational. That difference between referential content and cognitive content is entirely due to the fact that *cognitive content is transparent to the subject, while referential content is not*.

The appeal to cognitive content, distinguished from referential content, to account for Frege cases only makes sense on the assumption that cognitive content is transparent. The subject's rationality is said to be preserved because, internally, there is no contradiction, even though there is one at the level of referential content. According to RCT, rationality consists in the avoidance of internal contradictions — contradictions at the level of cognitive content. But to avoid contradictions at that level the subject 'must be capable of reflecting critically upon his or her own thoughts ; that sort of reflexive control over one's thoughts is possible only if they are transparently accessible' (Recanati 2012 : 217). If modes of presentation themselves were not transparent, there would be no reason to move from pure referential talk to mode of presentation talk in the explanation of rational behaviour.

The transparency of thoughts follows from the role they play in our practice of psychological explanation and rational evaluation of action (including mental and linguistic action). We account for the subject's behaviour, and assess her rationality, by appealing to the contents of her intentional states — what she believes, what she desires, what she intends, etc. These contents are supposed to be transparent to the subject.[2] Without the transparency of thoughts, Boghossian says, the practice of psychological explanation and appraisal would hardly make sense :

> we . . . ascribe thoughts to a person . . . for two related purposes; one the one hand, to enable assessments of his rationality and, on the other, to explain his behavior. As these matters are currently conceived, a thought must be epistemically transparent if it is to play these roles. Without transparency, our conceptions of rationality and rational explanation yield absurd results. We manifest recognition of this fact by barring de re thoughts [= contents referentially individuated] —thoughts which intuitively lack epistemic transparency—from figuring in assessments of rationality and psychological explanation. (Boghossian 1994 : 39)

[2] As Laura Schroeter writes, 'Most philosophers of mind accept the . . . thesis that you have transparent access to the contents of your own thoughts: provided that you're minimally rational, you simply cannot mistake one conceptual content for another' (Schroeter 2007 : 597).

It is because referential content is not transparent that it cannot play the role cognitive content plays in the explanation and assessment of behaviour. Hence the need for a distinction between two notions or dimensions of content (McGinn 1982), more or less along Frege's lines.

3. Modes of presentation as vehicles

So far we haven't said what modes of presentation are. The only thing we know is that they are transparent to the subject (in contrast to the entities they are modes of presentation of), and that they obey 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)

This very minimal characterization of modes of presentation leaves considerable latitude to the theorist. In particular, one does not have to opt for a 'semantic' or full-blooded construal of modes of presentation. One may opt for a 'syntactic' or austere construal instead.

On what I call the semantic, full-blooded construal, initiated by Frege, a linguistic expression is associated with a 'sense' which mediates between the expression and its reference. The sense is *a set of conditions which an item in the world has to satisfy in order to count as the reference of the expression*. Equivalently, the sense of an expression is *a collection of things known about the reference* — the reference being the entity which fits the body of knowledge in question. Competent language users know what the conditions are, even if they do not know which entity in the world fits those conditions. This captures the difference between sense and reference vis à vis transparency.

There are two major objections to that 'descriptivist' construal of modes of presentation. They are well-known, and I will be brief.

First, there is Kripke's argument, or family of arguments, from ignorance and error (Kripke 1980). A linguistically competent subject may not possess the sort of identifying knowledge of the reference which that conception takes to be required to grasp the 'sense' of an expression ; instead, she may possess only 'unsteady and confused notions', as Locke puts it (Locke 1690 : IV, X, §4). Such notions are insufficient to determine the reference. Or, worse, the subject may be radically mistaken concerning the nature and properties of the reference. The possibility of such mistakes about the reference shows that the reference cannot be what fits the subject's conception (otherwise the subject could not make mistakes).

The second major objection to the semantic construal is the buck-passing objection. If the reference of an expression is determined by a set of conditions mentally represented by the linguistically competent subject, and constituting its sense, then the relevant mental representations in the subject's mind will have to refer to the conditions in question. But if the reference of these mental representations is determined in the same way, via a set of

conditions, then an infinite regress is launched. 'Sooner or later', Pylyshyn says, 'the regress of specifying concepts in terms of other concepts has to bottom out' (Pylyshyn 2001 : 129). As Devitt puts it,

> There must be some representations whose referential properties are not parasitic on those of others, else language as a whole is cut loose from the world. Description theories pass the referential buck, but the buck must stop somewhere. It stops with theories (…) that explain reference in terms of direct relations to reality (Devitt 2014 : 477).

This takes us to the alternative, syntactic approach to modes of presentation, an approach advocated by authors such as Fodor (1998) and Sainsbury and Tye (2012).

Speaking about Frege cases (e.g. 'Cicero was Roman but Tully was not Roman'), Kamp writes :

> The two beliefs are, externally speaking, about the same object, and as such they are inconsistent. Yet in an internal sense they are not about the same thing, and thus they do not appear inconsistent to the believer himself (Kamp 1985: 250)

Kamp himself unambiguously opts for a syntactic construal of cognitive or 'internal' content. For him the difference between internal and external inconsistencies is this : if the same *object representation* is deployed twice in the subject's thought ('A is F and A is not F'), there is internal inconsistency ; if different object representations (representing the same object) are deployed ('A is F and B is not F'), the inconsistency is external. When the mental representation is 'A is F and A is not F' the subject knows that the same entity is referred to twice.

The syntactic approach revives the basic equation at the level of *mental* representations. The semantic content of a mental representation is said to be its reference, period. What plays the role of mode of presentation is *the mental representation itself*, qua syntactic entity, rather than an additional level of semantic content mediating between the expression and its reference. The subject who believes that Cicero was Roman but that Tully was not entertains conflicting attitudes towards the same referential content because that content is apprehended via *distinct mental representations* (involving two distinct mental files : a 'Cicero' file and a 'Tully' file). So, as far as semantics is concerned, we need only two things (the representation and its reference) rather than three (the representation, its sense, and its reference).

Is the same move available for linguistic expressions ? Can we say that we need only two things : the reference of the expression (its semantic content), and *the expression itself* serving as mode of presentation and accounting for Frege cases ? This is what Mates' cases may seem to suggest : there being two different *words* (eg 'psychiatrist'/'alienist', or 'Greek'/'Hellene') is sufficient to make Frege cases possible (Mates 1950). Note, however, that the existence of distinct words is not *necessary* to generate Frege cases. Even if there is a single word in the language, e.g. 'Paderewski' as the proper name of the Polish citizen who was well-known both as a politician and as a pianist, Frege cases will still be possible if the subject associates distinct mental files with that name (thinking there are two distinct Paderewskis, Paderewski the musician and Paderewski the politician).[3] This shows that it is

[3] The 'Paderewski' example is discussed in Kripke 1979.

the associated mental representation, not the linguistic expression itself, which matters to cognitive content. Indeed, the existence of distinct words is not even sufficient to generate Frege cases, contrary to what Mates' cases superficially suggest : the subject may treat 'Cicero' and 'Tully' as two names of the same individual, and *associate both of them with the same mental file for Cicero/Tully*. No Frege case can be generated in these circumstances. Anaphora provides a striking example in which two distinct linguistic expressions are associated with the same mental representation, in such a way that their cognitive content is the same. In *Mental Files in Flux* (Recanati 2016) I discuss the following example :

> I saw John$_i$ the other day. The bastard$_i$ did not greet me.

The mental file associated with the proper name 'John' in the first sentence is redeployed in association with the anaphoric description 'the bastard' in the subsequent sentence. As a result no Frege case is possible : a rational and linguistic competent subject cannot ascribe contradictory properties to John and to 'the bastard', since, simply in virtue of understanding the discourse, she knows they are one and the same individual.

Let us take stock. Two distinct linguistic expressions may be associated with the same mental representation (as in anaphora), in which case their cognitive content is the same ; or two occurrences of the same linguistic expression type may be associated with two distinct mental representations (as in the Paderewski example), in which case the cognitive content will vary across occurrences of the same linguistic expression. This suggests that, for linguistic expressions, we can stay somewhat closer to the Fregean position and maintain that we need three things rather than two. The three things are : the linguistic expression, the associated mental representation, and the reference of the mental representation, which the linguistic representation inherits. On this view, what plays the role of mode of presentation for a linguistic expression is neither a sense construed as an aspect of its semantic content, nor the linguistic expression itself, but *the associated mental representation*, construed as a syntactic entity (a mental file, in the case of singular terms).

In postulating these three levels (linguistic representation, mental representation, and reference) we follow Chastain's early piece of advice :

> A theory of singular reference will have to be combined with a systematic account of certain internal states of the speaker—his thoughts, beliefs, perceptions, memories, and so on—which are, so to speak, the intermediate links connecting the singular terms he utters with their referents out in the world. These intermediaries can themselves be understood only if we treat them as being quasi-linguistic in structure and content . . . and as containing elements analogous to singular terms which can be referentially connected with things in the world… (Chastain 1975 : 197)

On the mental file picture I have developed (Recanati 2012 and 2016), a singular term such as the name 'Cicero' has a reference (Cicero), which is its semantic content. It inherits that reference from the mental file it is associated with (the 'Cicero' file), which file itself refers to Cicero in virtue of informational connections to Cicero. (On that picture, the reference of a file is not determined by the information it contains, but by the informational connections it exploits.) 'Cicero was Roman' and 'Tully was Roman' carry distinct cognitive contents for the subject to the extent that the names are associated with distinct coreferential files in the subject's mind, a 'Cicero' file and a 'Tully' file. As we have seen, that need not be the case :

the subject may treat 'Cicero' and 'Tully' as two names of the same individual, and associate both of them with the same mental file for Cicero/Tully.

4. The transparency of the vehicle

Modes of presentation, and the thoughts of which they are constituents, are transparent to the subject. What does that mean, if modes of presentation (and thoughts) are understood syntactically as mental representations ? Let us, again, focus on the case of singular terms. In a Frege case such as that involving Cicero and Tully, the proper names 'Cicero' and 'Tully' are associated with distinct mental files, because the subject does not know that the two names corefer. She takes 'Tully' to refer to an individual distinct from Cicero. That is why she can, without irrationality, ascribe contradictory properties to Tully and Cicero, by storing the predicate *being Roman* in the 'Cicero' file, and the predicate *not-being-Roman* in the 'Tully' file. No contradiction arises within any of the two files, so the subject's rationality is not impugned. If, apprised of the identity, the subject associated a single Cicero/Tully file with the two names, the contradiction between the two predications would become apparent because it would be internal to the file in question. Of course, this presupposes that the subject knows which predicates are bound to which files ; and, more fundamentally, that *the subject knows when they are deploying the same file twice, and when they are deploying two distinct files*. The transparency of cognitive content now becomes the transparency of the vehicle.

Let us assume that the subject, indeed, knows when they are deploying the same file and when they are deploying distinct files. It follows that there are two types of coreference. We saw earlier that, because reference is not transparent, coreference is not transparent either : it is possible for the subject to refer to the same individual twice (as in Frege cases) without knowing that the two acts of reference target the same individual (§2). That happens whenever the subject associates two distinct yet coreferential files with two token singular terms. The subject may not know that the two files corefer, so her ascribing contradictory properties to the same object (the referent of the two files) does not threaten her rationality. But if the subject deploys the same file twice, in association with both of the singular terms, *she is bound to know that she is referring to the same entity twice* (assuming she succeeds in referring). Here the transparency of cognitive content, understood as the transparency of the vehicle, entails the transparency of coreference. So the two types of coreference I announced are : *de facto* coreference, which is opaque (the subject may not know that the two coreferential expressions corefer), and *de jure* coreference, which is transparent (the subject is bound to know that the two coreferential expressions corefer). In cases of *de facto* coreference the subject associates two distinct yet coreferential files with two token expressions ; in cases of *de jure* coreference the subject associates the same mental file with two token expressions.

The modal formulations ('may not know…', 'is bound to know') are necessary because there can be *de facto* coreference even if the subject *actually knows* that the two coreferential expressions corefer. Identity statements are a case in point. Consider the following example, adapted from Higginbotham (1985) :

Peter : Who is that guy ?
Mary : He put on John's coat, so, believe me, he is John.

The pronoun 'he' and the name 'John' in Mary's utterance are associated with two distinct mental files, a demonstrative file for the guy both interlocutors are looking at, and their regular file for John. Mary happens to know that the two files corefer, and her identity statement is meant to convey that truth to the hearer ; but the knowledge in question is contingent in the sense that someone might fully understand the statement she makes without accepting it and taking the two terms to corefer. For example, Peter might respond :

> You're wrong. He put on John's coat, but he is not John.

In cases of coreference *de jure*, however, no one can *understand* the discourse — grasp its cognitive content — while doubting that there is coreference or even wondering whether there is.[4] The coreference relation is entailed by the (transparent) identity of the mental representations respectively associated with the two singular terms. As Kit Fine puts it,

> A good test of when an object is represented as the same[5] 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*. 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 ; emphasis mine).

5. Coreference *de jure* as transparent coreference

I said that, when the same mental file is deployed twice, the subject is bound to know that there is coreference. In such cases, (i) there is coreference at the level of referential content (the same entity is referred to twice), but (ii) in addition there is a matching relation holding at the level of cognitive content between constituents of the mental representation (the same mental file occurs twice). Being thus reflected at the level of cognitive content, the coreference relation holding at the level of referential content is made transparent to the subject. This is coreference *de jure*. (In cases of coreference *de facto*, the coreference relation holding at the level of referential content is not reflected at the level of cognitive content : two distinct mental files are deployed, which happen to refer to the same entity.)

But there are (alleged) counterexamples to the claim that coreference is made transparent whenever the same file is deployed twice. These counterexamples fall in two categories : cases of emptiness, in which the mental file which is deployed twice fails to refer ; and cases of confusion, in which the file which is deployed twice tracks different entities on its two deployments. In such cases it cannot be said that, thanks to the matching

---

[4] That is what we saw in the previous section in connection with the 'bastard' example ('I saw John$_i$ the other day ; the bastard$_i$ did not greet me'). Because of the anaphoric link, whoever understands the discourse knows that 'John' and 'the bastard' are bound to corefer.

[5] 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 'he is John' or 'Cicero is Tully' is said to represent the referent of the two singular terms as *being* the same.

relation at the level of cognitive content, the subject knows that the coreference relation obtains at the level of referential content ; for the coreference relation does *not* obtain.

The following examples illustrate situations in which the same mental file is deployed twice, in association with two distinct occurrences of singular terms, yet (it is argued) the singular terms in question do not corefer :

(1) Look at that dagger$_i$! It$_i$ has blood on it.
(2) The guy in the corner$_i$ is a bit strange. He$_i$ is so silent. Look, he$_i$ is staring at you now !

In (1) the dagger is hallucinated and does not exist. The referring expressions 'that dagger' and the anaphoric pronoun 'it' therefore do not refer, whence it follows that they do not corefer (since coreference entails reference). In (2) the subject deploys a demonstrative file for the guy in the corner, and redeploys that file twice in association with the two subsequent occurrences of the anaphoric pronoun 'he'. But let us suppose that, between the two occurrences of 'he', an evil genius has magically replaced the strange and silent guy in the corner by a Doppelgänger who happens to be staring at the hearer, and whose behaviour the speaker attempts to report by saying 'Look, he is staring at you'. In this case both the description 'the guy in the corner' and the first occurrence of 'he' refer to the silent guy in the corner, while the second occurrence of 'he' attempts to refer to the Doppelgänger, under the mistaken presupposition that he is identical to the previously mentioned guy in the corner. This is a case of confusion in which, through the deployment of a single file, the subject attempts to refer to two distinct individuals.

Cases of that sort have motivated the abandonment of the idea of transparent coreference. According to Krista Lawlor (2010), the subject who deploys the same mental file twice presupposes or takes for granted that she refers to the same entity twice ; but that need not be actually the case, so this is not 'transparent coreference'. Since the two terms may not be coreferential, Lawlor rejects the claim that the subject who deploys the same file twice thereby *knows* that there is coreference. Likewise, Kit Fine distinguishes between strict coreference and putative coreference (Fine 2010). In strict coreference there is both coreference at the level of referential content and a matching relation at the level of cognitive content ; but the matching relation at the level of cognitive content may obtain in the absence of actual coreference, as in the counterexamples. That relation, by itself, only yields *putative* knowledge of coreference. Again, this is not 'transparent coreference'.

I do not find these counterexamples convincing. To deal with the empty cases, it is sufficient follow a suggestion due to Fine himself :

> 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)

In other words, when a file that fails to refer is deployed twice, as in example (1), what the subject knows is that she refers to the same entity twice *provided she succeeds in referring*. She knows that the two terms M and N associated with that file refer to the same thing *if*

they refer at all. It is this weaker relation of (conditional) coreference which is transparently known to obtain, whether the terms are empty or not.

What about the confusion cases ? They are trickier and well worth investigating (see Recanati 2016) but, appearances notwithstanding, they are irrelevant to the issue at stake. Such cases only occur when we consider deployments of a given file *at different times*. In (2) the file for the guy in the corner is deployed first at a time t_1 when the interlocutors are looking at him, but then it is redeployed at a later time t_2 when (after the intervention of the evil genius) what the interlocutors are looking at is the Doppelgänger. This temporal difference is problematic, for what rationality requires is *synchronic* transparency. A fully rational subject must be able to detect contradictions within a given train of thought, *at a particular time*. A problem arises for transparency only if, for the same subject at the same time, distinct occurrences/deployments of the same file do not refer (or conditionally refer) to the same entity. (Note that a synchronic condition is explicitly built into Frege's constraint on modes of presentation.)

Of course, language is processed in time : in oral speech, distinct singular term occurrences are bound to be pronounced one after the other. But that is not the sort of temporal difference that matters. What matters is whether the files associated with two singular term occurrences are the same file in the strong sense, or merely in the weak sense of being (distinct) temporal stages of the same 'dynamic' file. Being associated with the same file in the weak, dynamic sense is not sufficient to yield transparent coreference, and I am not claiming that it is. So diachronic examples like (2) are not counterexamples to my claim that deployments of the same file (in the strong sense) yield transparent knowledge of coreference. [6] Let me explain.

Mental files develop, grow, merge, and split. This gives rise to a notion of dynamic file: a continuant made up of successive file-stages. Because confusion can always arise as information is collected across time, It is always possible for one stage to refer, while the next stage fails to refer due to confusion. The fact that two nonsynchronous file deployments, such as those associated with the two occurrences of 'he' in (2), are deployments of the same dynamic file is therefore not sufficient for the subject to know that the coreference relation obtains between these occurrences. That is so *even if we define the coreference relation in the manner suggested by Fine*, i.e. as

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

In example (2), the subject starts by referring (twice) to the guy in the corner, but the third deployment of the same dynamic file fails to refer because it now tracks two objects at the same time, the guy in the corner (through the anaphoric link to the initial acts of reference), *and* the Doppelgänger (through the perceptual link to him). Condition (3) is not satisfied here : even though they are associated with the same dynamic file, one of the singular terms (the description or the first occurrence of the pronoun) refers to some object *x* (the guy in

[6] A reviewer suggests a simpler way of disposing of alleged counterexamples like (2). The second occurrence of the pronoun may be construed as referring to the same individual as the definite description and the first occurrence of the pronoun. On that understanding, there *is* transparent coreference to the initial 'guy in the corner' throughout the discourse, but the clause 'he is staring at you now' says something false since the guy staring at the addressee is not the referent (the initial individual) but the Doppelgänger.

the corner), while the second occurrence of the pronoun arguably fails to refer because it tracks both *x* and some other object *y* (the Doppelgänger).

Even though language is processed in time, the mental files that are associated with distinct constituents in a single utterance can be thought of as synchronously deployed, because they are *codeployed in the thought which is the output of the interpretation process*. We should therefore only be concerned with *synchronous* deployments of the same file, i.e. with deployments of *the same file-stage* (rather than deployments of different temporal stages of the same dynamic file, as in the alleged counterexamples to transparency based on confusion). Such synchronous deployments of the same file do make the (conditional) coreference relation holding at the level of referential content transparent to the subject ; that's what I am claiming.[7]

6. Further alleged counterexamples

Some people claim that there can be failure of transparency at the synchronic level, i.e. there can be cases in which the same file is deployed twice (at a given time) yet the weak coreference condition (3) is not satisfied. I find none of the supporting examples convincing, and I will end this paper by briefly discussing two of them.[8]

Boghossian imagines a synchronous train of thought involving referentially divergent deployments of the same file. His example is a variant of Burge's 'slow switching' case in which, unbeknownst to him, a subject is switched to Twin-Earth and stays there for a very long time.[9] When that subject, Peter, arrives on Twin-Earth, his water thoughts refer to water and when, pointing to lakes and rivers on Twin-Earth, he says (or thinks), 'This is water', he is wrong: the propositions he entertains are false (since the stuff is actually twater, not water). But after many years on Twin-Earth the subject's water thoughts, or some of them at least, will no longer be water thoughts: they will be regular twater thoughts. In other words, their referential content will have switched even though, as Burge puts it, 'the introspection remain[s] the same' (Burge 1988 : 652). This is a diachronic failure of transparency. To make the failure synchronic, hence problematic, Boghossian imagines

---

[7] In *Mental files in Flux* (Recanati 2016) I make a further claim : that it is possible to define a notion of coreference that is even weaker than that corresponding to Fine's 'universal' definition (itself weaker than that corresponding to the existential definition) ; and that, using that weaker notion, we can construct a correspondingly weak notion of coreference *de jure* (still construed as knowledge of coreference), which covers even diachronic deployments of the same dynamic file.

[8] A reviewer for this journal mentions 'another apparent counterexample' where 'due to whatever reasons (maybe, a general belief that people may sometimes be replaced by impostors), a person may believe at t_1 that a certain individual, whom she has faced at t_0, is not the same as another individual, whom she faces at t_1, notwithstanding the fact that she still associates at t_1 precisely the very same characteristics (possibly also the very same name) to the allegedly different 'individuals'. In point of fact, however, there is just one individual over there that such a person faces. This seems to show that mobilizing the same mental file twice (…) is not enough in order to have *de jure* coreference.' Here, however, contrary to what the reviewer assumes, I would deny that the same mental file is deployed twice.

[9] On 'Twin-Earth', see Putnam 1975 ; on 'slow switching', see Burge 1988.

that Peter has both pre-switch memories of a striking event involving water, and 'undated general thoughts' about 'water' which (because the subject has been on Twin-Earth for a long time) refer to twater. Boghossian assumes that

> those tokens of 'water' occurring in memories , and in beliefs about the past based upon them, will retain their Earthly interpretations, despite being tokened on Twin-Earth. Such thoughts, unlike, for instance, beliefs with undated general contents, or thoughts about one's present surroundings, are caused and sustained by previous perceptions long gone. In the normal case, they owe little, if anything, to current perceptions and cognitive transactions with one's environment. From a purely intuitive standpoint, they would be expected to retain their Earthly interpretations, despite the admitted shift in their syntactic cousins. (Boghossian 1994 : 38)

Based on that controversial assumption[10] Boghossian concludes :

> In the situation described, Peter's externally individuated thought tokens are not epistemically transparent to him. In particular, Peter's language of thought contains tokens expressions that possess different semantic values, despite being of the same syntactic type.(...) From the inside, however, there will be no indication of this : as far as Peter is concerned, they will appear to express precisely the same contents. (Boghossian 1994 : 39)

Boghossian's thought experiment is meant to show that there is a conflict between transparency and Burge's externalist doctrine. Since Boghossian thinks transparency is not negotiable, the conflict in question is supposed to provide a reason to reject externalism. But I don't think Boghossian succeeds in showing that there is a conflict. There are several alternative analyses of the example, which do *not* posit any referential divergence in synchrony (see Recanati 2012, chapters 10-11 for discussion). For example, we can say that Peter does retain memories of past events, but that these memories are *reinterpreted* in the light of the (shifted) concepts currently available to him (Sainsbury and Tye 2012).

A related but more mundane example supposedly showing the possibility of synchronous failures of transparency involves a subject touching a certain object while seeing an object which he (wrongly) presupposes to be the same object he is touching. The subject makes the following inference :

> This is rough, this is red, therefore, something is red and rough.

The first demonstrative 'this' is supposed to refer to the touched object, while the second demonstrative refers to the seen object. The conclusion follows only on the assumption that they are the same, and this is indeed presupposed by the subject who deploys the same file thoughout. There is failure of synchronic transparency to the extent that, on the first deployment, the file refers to the touched object, while on the second deployment it refers to something else, namely the seen object. The weak coreference condition (3) is not satisfied, on that understanding of the example.[11]

[10] For a critique of Boghossian's assumption see Burge 1998 : 367.

[11] I owe this objection to Rachel Goodman.

But that is not how *I* would analyse the example. What I would say is this. Because of the presupposition that one is seeing and touching the same object, a single file (based on both vision and touch) is deployed throughout ; and that file fails to refer (on both of its occurrences) because it tracks two distinct objects at the same time. On that type of analysis (which can also be provided for Boghossian-like cases, as I argue in Recanati 2012 : 131-32), the weak coreference condition (3) *is* satisfied.

References

Boghossian, P. (1994) The Transparency of Mental Content. *Philosophical Perspectives* 8: 33–50.

Burge, T. (1988) Individualism and Self-Knowledge. *Journal of Philosophy* 85: 649–65.

Burge, T. (1998) Memory and Self Knowledge. In P. Ludlow and N. Martin (eds.) *Externalism and Self Knowledge*, 351–70. Stanford: CSLI.

Chastain, C. (1975) Reference and Context. In K. Gunderson (ed.) *Language, Mind, and Knowledge*, 194-269. Minneapolis : University of Minnesota Press.

Devitt, M. (2014) Lest Auld Acquaintance Be Forgot. *Mind & Language* 29: 475-484.

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

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

Fodor, Jerry (1998) *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press.

Frege, G. (1967) *Kleine Schriften*, ed. I. Angelelli. Hildesheim : Olms.

Higginbotham, J. (1985) On Semantics. *Linguistic Inquiry* 16 : 547-93.

Kamp, H. (1985) Context, thought and communication. *Proceedings of the Aristotelian Society* 85 : 239–261.

Kripke, S. (1979) A Puzzle about Belief. In A. Margalit (ed.) *Meaning and Use*, 239–83. Dordrecht: Reidel.

Kripke, S. (1980) *Naming and Necessity*. Oxford : Blackwell.

Lawlor, K. (2010) Varieties of Coreference. *Philosophy and Phenomenological Research* 81 : 485-501.

Locke, J. (1690) *An Essay Concerning Humane Understanding*. London : Thomas Basset.

McGinn, C. (1982) The Structure of Content. In A. Woodfield (ed.) *Thought and Object*, 207-58. Oxford : Clarendon Press.

Mates, B. (1950) Synonymity. *University of California Publications in Philosophy* 25 : 201-26.

Putnam, H. (1975) The Meaning of 'Meaning'. In his *Mind, Language and Reality : Philosophical Papers, vol. 2*, 215-71. Cambridge : Cambridge University Press.

Pylyshyn, Z. (2001) Visual Indexes, Preconceptual Objects, and Situated Vision. *Cognition* 80: 127-158.

Recanati, F. (2012) *Mental Files*. Oxford: Oxford University Press.

Recanati, F. (2016) *Mental Files in Flux*. Oxford: Oxford University Press.

Sainsbury, M. and Tye, M. (2012) *Seven Puzzles of Thought* . 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