



**HAL**  
open science

## Mental Files

François Recanati

► **To cite this version:**

François Recanati. Mental Files. The Cambridge Handbook of the Philosophy of Language, 1, Cambridge University Press, pp.535-547, 2021, 10.1017/9781108698283.030 . ijn\_03470852

**HAL Id: ijn\_03470852**

**[https://jeannicod.ccsd.cnrs.fr/ijn\\_03470852](https://jeannicod.ccsd.cnrs.fr/ijn_03470852)**

Submitted on 8 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Mental files

F. Recanati

Collège de France

To appear in  
*The Cambridge Handbook of  
the Philosophy of Language*  
(Piotr Stalmaszczyk ed.)

### 1 Mental files as descriptive modes of presentation

A rational subject, S, may take different (and possibly conflicting) attitudes towards the judgment that a given individual is *F* — for example, she may reject it as false or accept it as true — depending on how that individual is presented. She may think that *Emile Ajar* is a genius, while thinking that *Romain Gary* is a has-been (hence, not, or no longer, a genius). Since *Emile Ajar is Romain Gary*, our subject S ascribes to the same individual the property of being and not being a genius. Is that irrational of her? No, of course, for she does not *know* that Ajar is Gary.

The proper names ‘Ajar’ and ‘Gary’ refer to the same individual, and, according to a widespread view (the Direct Reference thesis or DR for short), the reference of a proper name is its semantic value. From DR it follows that ‘Ajar is a genius’ and ‘Gary is a genius’ express the same proposition (to the effect that Gary/Ajar is a genius), a proposition which the subject both accepts and rejects when she simultaneously holds that Ajar is a genius while Gary isn’t. If the subject thus contradicts herself by believing and disbelieving one and the same content, how can we account for her rationality? DR also entails, or seems to entail, that ‘Ajar is Gary’ and ‘Ajar is Ajar’ express the same trivially true proposition (to the effect that Gary/Ajar is self-identical). What is it, then, that the subject learns when she hears (at Gary’s death) that Ajar was Gary? Why is ‘Ajar is Gary’ informative while ‘Ajar is Ajar’ is not, if they express the same proposition?

To answer these questions, Frege rejected DR, in two steps. First, he assigned to (occurrences of) linguistic expressions a ‘sense’, in addition to their reference. The sense is

the way the reference of the expression is presented — something that, like the reference, may depend upon the context of use. Second, he took the sense of an expression, rather than its reference, to be its contribution to the proposition (or, rather, the ‘thought’) expressed by the sentence in which it occurs.

The idea is that ‘Ajar’ and ‘Gary’ refer to the same individual, but present him in different ways. The referent of ‘Ajar’ is supposed to have a certain number of characteristic properties : bearing the name ‘Emile Ajar’, being the author of *La Vie devant soi* and *L’Angoisse du roi Salomon*, having won the Goncourt Prize in 1975. The referent of ‘Romain Gary’ is supposed to have a different set of properties : being a war hero and a former French Ambassador, having been Jean Seberg’s husband, being named ‘Romain Gary’, being the author of many books, including *la Promesse de l’Aube*, having won the Goncourt Prize in 1956. These ‘two’ individuals turn out to be the same : late in his career, the war hero and former diplomat Romain Gary, author of famous books including one which won the Goncourt Prize, decided to use the name ‘Emile Ajar’ as a pseudonym and, under that fictitious name, wrote a series of other acclaimed books, and thus managed to win the Goncourt Prize for the second time.

According to Frege, what an expression like ‘Ajar’ or ‘Gary’ contributes to the thought expressed by the sentence in which it occurs is its sense, not its reference. So ‘Ajar is a genius’ and ‘Gary is a genius’ do *not* express the same thought : the subject’s rationality is thus preserved. Likewise, an identity statement like ‘Ajar is Gary’ is informative, in contrast to the truistic ‘Ajar is Ajar’, because the senses associated with the two names are different. Thanks to these different senses, the subject who hears that Ajar is Gary learns a number of contingent facts, e.g. that the author of *La Vie devant soi* is the author of *La Promesse de l’Aube*. In contrast, the subject who hears that Ajar is Ajar does not learn anything. The two identity statements express different thoughts, one trivial and the other not.

Now what *are* senses or modes of presentation ? Frege himself thought of them as *descriptions* of the referent. When we think about an object, we think of it as ‘the *F*’, where *F* is some uniquely-identifying property of the referent, e.g. ‘the author of *La Vie devant soi*’ or ‘the author of *La Promesse de l’Aube*’. Because the knowledge of the reference associated with a proper name typically involves a rich body of information rather than a single identifying description, a variant of the Fregean approach has emphasized the fact that descriptions come in *clusters*. The ‘cluster’ idea is to be found in the work of Wittgenstein and Searle, but also in that of Strawson and Grice. A cluster of descriptions can be thought of

as a *mental file* or *dossier* about an individual, involving many distinct pieces of information, as in the above examples.

Is a cluster or dossier the same thing as a long, conjunctive description ('the  $x$  such that  $x$  was Jean Seberg's husband &  $x$  wrote *La Promesse de l'Aube* &  $x$  was named 'Romain Gary' &...')? Not quite. For one thing, different descriptions in the cluster can be given different weights, and the weights themselves can be modulated in context. For another, a cluster can persist as 'the same cluster' even though some descriptions in the cluster are replaced by others. The reference of the dossier is the item 'which satisfies the majority of, or each member of a specially favoured subset of, the descriptions in the dossier' (Grice 1969 : 142). A particular description  $\delta$  will be expelled from the dossier if the subject 'comes to think that  $\delta$  is a misfit in the [subject's] dossier, i.e. that  $\delta$  is not, after all, satisfied by the same item as that which satisfies' the rest of the descriptions in the dossier (*id.*). As a result, there is some vagueness inherent in the notion of cluster, a vagueness which distinguishes clusters from conjunctive descriptions. Wittgenstein emphasized that vagueness in the well-known passage cited by Kripke (1980 : 31) :

If one says 'Moses did not exist', this may mean various things. It may mean: the Israelites did not have a single leader when they withdrew from Egypt— or : their leader was not called Moses — or : there cannot have been anyone who accomplished all that the Bible relates of Moses — ... But when I make a statement about Moses, am I always ready to substitute some *one* of those descriptions for 'Moses'? I shall perhaps say: by 'Moses' I understand the man who did what the Bible relates of Moses, or at any rate, a good deal of it. But how much? Have I decided how much must be proved false for me to give up my proposition as false? Has the name 'Moses' got a fixed and unequivocal use for me in all possible cases? (Wittgenstein 1953 : §79)

## 2. Mental files as nondescriptive modes of presentation

Whatever the details, Frege's descriptivist approach has largely fallen into disrepute. Several *anti-descriptivist arguments* have convinced most philosophers — though not all of them<sup>1</sup> — that an alternative was needed. Let me mention three influential arguments :

---

<sup>1</sup> There are neo-descriptivists who try to rebut the standard objections to Fregean descriptivism. See e.g. Sosa (1995) and Jackson (1998).

- *The argument from perception*

When we perceive an object and have a thought about it, the object the thought is about is the object the perception is about ; and that, arguably, is *not* determined by properties the subject takes the referent to have (Pylyshyn 2007). Often, we are unable to properly describe the object that is given to us in experience : we don't know what it is, yet that does not prevent us from referring to it directly (without conceptual mediation) and e.g. wondering what *it* can be (Dretske 1988 : 73) In such cases, even though we are unable to *conceptually articulate* what our thought is about, 'Frege cases' are still possible. I may be perceptually related (through distinct sense modalities, say) to what I take to be two objects, which happen to be one and the same object.<sup>2</sup>

- *The argument from indexicals*

Indexicals like 'I', 'here' and 'now' present their reference in a manner that is essentially perspectival and *cannot be captured by means of objective, non-indexical descriptions*. For any indexical  $\alpha$  and non-indexical description 'the F', it is always possible for the subject to doubt, or to wonder, whether  $\alpha$  is the F (Castañeda 1989). The impossibility of substituting a descriptive paraphrase for an indexical without altering the mode of presentation is the celebrated 'problem of the essential indexical' (Perry 1979), which has given rise to the rich body of literature about *de se* thoughts (Lewis 1979, Perry 1993, Stalnaker 2008, Carpintero and Torre 2016).

- *The argument from ignorance and error*<sup>3</sup>

In 'Frege cases' involving proper names ('Ajar'/'Gary', 'Hesperus'/'Phosphorus', 'Cicero'/'Tully' etc.), it is often the case that the subject, although able to describe the referent, can only provide *indefinite* descriptions ('a famous Roman orator'), i.e. descriptions which, because they are multiply satisfied, cannot determine the reference of the name. Such ignorant subjects can still refer by means of proper names, and this shows that the reference by a name is not determined by the descriptions in the mind

---

<sup>2</sup> To be sure, as the neo-descriptivists mentioned in footnote 2 would readily point out, the reference-fixing description in the mind of the subject might be something like 'what I am now seeing' or 'what I am now touching' ; but this supposes, on the part of the subject, *reflective abilities* the exercise of which is arguably not necessary to suffer from identity confusions of the type which Frege cases illustrate.

<sup>3</sup> The label 'argument from ignorance and error' comes from Michael Devitt (see e.g. Devitt and Sterelny 1999: 54-64). The label 'Frege cases' (used in the previous footnote and throughout this chapter) comes from Jerry Fodor (1994).

of the language user (Kripke 1980). Besides being ignorant, Kripke points out, subjects can also be mistaken. Even if the subject is able to provide a *definite* description of what she takes herself to be referring to, the descriptive condition she cites may fail to be satisfied by a unique object, as a matter of fact (even though, by using the name, the subject manages to refer to a determinate individual). Or, assuming the description the subject can provide is both definite and uniquely satisfied, *the satisfier need not be the referent of the name*, as in Kripke's 'Gödel'/'Schmidt' example. Again, this shows that the reference of a name is not determined by the descriptions in the mind of the language user.

The three types of objection point in the same direction. On the descriptivist picture, reference is determined *satisfactionally* : the referent is what satisfies the (majority of) descriptions in the subject's mind. But in the three type of case we talked about (perceptual reference, indexical reference, and reference via names) what seems to be crucial are the subject's *relations* to the objects of thought : reference is determined relationally, not satisfactionally (Bach 1987). Demonstrative thought rests on a perceptual/attentional link to the object of thought, while indexical reference rests on various contextual relations — the relation each individual bears to himself or herself in the case of first person thought, the relation to the place one occupies in the case of *here*-thoughts, and the relation to the present time in the case of *now*-thoughts. For proper names, what matters is the historical chain of communication through which the subject is related to the reference of the name. In all three cases, what fixes the reference is the relation, not the satisfaction of descriptions in the thinker's mind.

The relational stance characteristic of the anti-descriptivist movement is entirely compatible with the mental file idea, which does not, by itself, entail descriptivism. Instead of saying, with Searle and Grice, that the reference of a dossier is what satisfies most or a specially favoured subset of the descriptions in the dossier, one can say the following. A mental file refers not to what satisfies the descriptions in the file (since the subject may be severely mistaken) but to what stands in the right relation to the file or, equivalently, to the subject who deploys the file. As Bach puts it,

Since the object of a descriptive thought is determined satisfactionally, the fact that the thought is of that object does not require any connection between thought and object. However, the object of a *de re* thought is determined relationally. For something to be

the object of a *de re* thought, it must stand in a certain kind of relation to that very thought. (Bach 1987 : 12)

The relevant relation is a relation serving as *information channel* : a relation to the object which makes it possible for the subject to gain information from that object, and thus to feed the file based on that relation (Recanati 2012). For example, a demonstrative file is based on the attentional relation, a relation that makes it possible to gain perceptual information from the object one is attending to. The information thus gained goes into the file. (A demonstrative file, thus understood, is a short-term file, which exists only as long as one is paying attention to the object. But long-term files too are based on relations to the reference, e.g. the ‘familiarity relation’ which holds whenever ‘multiple exposure to [an] object has created and maintained in the subject a disposition to recognize that object’ (Recanati 2012 : 71).)

Because they enable the subject to gain information from the object, the relations on which mental files are based, and which determine their reference, have been called *epistemically rewarding* (ER) relations.<sup>4</sup> They are also called *acquaintance relations* because they put the subject in epistemic contact with the object, by making information about it available (information which it is the function of the file to store).<sup>5</sup> Although very different from the attentional relation which underlies the deployment of demonstrative concepts, or the recognitional relation which underlies the deployment of recognitional concepts, the subject’s relation to himself or herself, namely identity, counts as an ER relation, because it makes it possible for the subject to gain information about herself through a particular information channel, namely ‘from inside’ (through proprioception, introspection, episodic memory, etc.) : *one can get information about an individual in this manner (viz. from inside) only if one is that individual*. The information thus gained goes into the subject’s SELF-file, a special file based on the identity relation (Perry 2012). Acknowledging the role of ER relations therefore provides a solution to the problem of the essential indexical : just like demonstrative concepts (‘that thing’), indexical concepts such as the concept of oneself, the

---

<sup>4</sup> See Recanati (2012). The notion is often misunderstood. Gaining information from the object (e.g. perceiving it) is not what I mean by ‘standing in some ER relation’ to it. ER relations are various relations to objects which, *given the subject’s cognitive equipment, normally make it possible* for the subject to gain information from the object. For example, *being in* a certain place is a relation to that place which normally makes it possible to gain information about it. The mental file corresponding to the word ‘here’ is based on that ER relation, which determines the reference of the file (namely, the place we are in — normally but not necessarily the same as the place we actually gain information from when we look around).

<sup>5</sup> A caveat: the notion of acquaintance at play here differs from Russell’s and from the way the term is currently used in epistemology. On acquaintance relations see Lewis 1999: 380-1.

concept of the present time ('now') or the concept of the place one occupies ('here') should be construed not as descriptive concepts ('the  $x$  such that  $Fx$ ') but as *mental files based on particular ER relations*. Only a subject in the right relation to the referent can think a thought involving such a concept. This captures the 'limited accessibility' feature supposedly characteristic of indexical thought (Recanati 2016: 110ff).

An important aspect of the framework is that, in order to play their part, mental files need to stand in relation to the entities they refer to, via the ER relation which the subject deploying the file stands in to these entities. To be so related mental files have to be construed as entities on their own right, that is, as *particulars* — 'concrete mental particulars', as Crimmins and Perry 1989 say (see also Crimmins 1992). The idea we came across in section 1, that clusters or dossiers have *identity conditions of their own* and are *not* to be individuated in terms of the descriptions in the dossier (since the descriptions in the dossier may change), was a first step in that direction.<sup>6</sup>

Construing mental files as particulars neatly solves the problem which ignorance of the reference raises for theories that posit senses or modes of presentation. Kit Fine formulates the problem as follows :

Surely one may learn something different upon being told 'Cicero = Tully' and upon being told 'Cicero = Cicero'. (...) It is hard to see how to account for this possible cognitive difference except in terms of a semantic difference. The main problem with the Fregean position is to say, in particular cases, what the difference in the meaning or sense of the names might plausibly be taken to be. Although there appear to be good theoretical reasons for thinking that there *must* be a difference, it seems hard to say in particular cases what it is. For as Kripke (1980) has pointed out, it seems possible for a speaker, or for speakers, to associate the same beliefs or information with two names, such as "Cicero" and "Tully." And if the information or beliefs are the same, then how can the sense be different? (Fine 2007 : 35)

A descriptivist might respond that the 'Cicero' file contains the metalinguistic piece of information 'called *Cicero*' while the Tully file contains 'called *Tully*'. This is enough of a

---

<sup>6</sup> How are mental files, considered as particulars, to be individuated? They cannot be individuated by ER relations, since the subject might stand in the same ER relation to the same object at the same time while being under the illusion there are two objects and therefore deploying two files about it (see the 'moth' example in Pagin 2013: 140). Arguably files are individuated by their *origin* (Sainsbury and Tye 2012): the origin of a file is a file-opening event, and in Pagin's 'moth' example, there are two such events, contrary to what Bonardi (forthcoming) wrongly assumes. See Tałasiewicz 2017: 30-32 for a similar idea.

difference in descriptive content between the two files. But, according to some theorists in the anti-descriptivist camp, there can be Frege cases even in the absence of any such metalinguistic difference (Pryor 2016). Let us assume this is right,<sup>7</sup> and ask Fine's question again : 'if the information or beliefs are the same, then how can the sense be different?' Construing mental files as particulars provides an obvious answer. If there are two distinct files, then there are two distinct modes of presentation, *even if the information in the two files is exactly the same* ('a Roman orator').

### 3. Mental files : syntax or semantics ?

I have just said that, if there are two numerically distinct files (possibly containing the same information), then there are two distinct modes of presentation. But are we still talking about *senses* ? Is the (purely numerical) difference between the two files construed as particulars a *semantic* difference ? Clearly not. The difference is 'syntactic' : it is a difference between the *mental representations* that are respectively deployed, rather than a difference in the content of the representations. The representations have the same content – they refer to the same individual and carry the same information about it – but they are distinct representations. Assuming a purely syntactic difference such as this is sufficient to generate Frege cases,<sup>8</sup> it was a mistake on Frege's part to argue from the possibility of Frege cases to the necessity of adding an extra layer of semantic content, that of 'sense', on top of the reference of expressions.

For years, Jerry Fodor has argued against Frege along these lines. There is only one layer of content, Fodor used to insist, and that is the reference. To account for Frege cases, the only thing we need, in addition to the content thus understood, is the *vehicle* or *bearer* of content : the representation through which reference is achieved.<sup>9</sup>

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

---

<sup>7</sup> This is actually controversial ; see Gray 2016..

<sup>8</sup> Again, this is controversial. Even if one accepts the idea of qualitatively identical but numerically distinct files, one can doubt that the existence of a pair of such files, as in Pagin's 'moth' example, could give rise to Frege cases. Indeed, as long as the files are indistinguishable, it is far from clear that the subject can deploy one (rather than the other) in a distinctive way to think - for example - of one of the alleged moths (rather than the other). In the absence of any distinctive element in the content of the files, it seems that the existence of two separate files only allows the subject to represent the cardinality of the set of moths he takes himself to perceive, that is, the fact that there are two, but not to relate distinctively to one of the alleged moths in thought. If this is true, then one cannot have a Frege case involving two numerically distinct but indistinguishable files.

<sup>9</sup> See Fodor (1998); see also Sainsbury and Tye (2012).

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)

If the same *object representation* was deployed twice in the subject's thought ('A is F and A is not F'), there would be inconsistency 'in an internal sense' ; but different object representations (representing the same object) are deployed ('A is F and B is not F'), so there is inconsistency only from an external point of view, and the subject's rationality is not impugned. What plays the role of internal component here is *the mental representation itself*, qua syntactic entity. 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, to account for Frege cases, 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).<sup>10</sup> This shows that it is the associated mental representation, not the linguistic expression itself, which matters. 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.<sup>11</sup>

---

<sup>10</sup> The 'Paderewski' example is discussed in Kripke 1979.

<sup>11</sup> Anaphora provides a striking example in which two distinct linguistic expressions are associated with the same mental representation. Consider the following example (from Recanati 2016) :

What I have just said suggests that, for linguistic expressions, we do need three things rather than two. But the three things are not, as for Frege, the expression, its sense, and its reference. The three things now are : the linguistic expression, *the associated mental representation*, and the reference of the mental representation, which the linguistic representation inherits. From this (psychologicistic and unFregean) perspective, 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, 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. Mental files thus construed are 'singular terms in the language of thought' (Recanati 2012: 182). They are syntactic entities.

Still, this conclusion can be mitigated by noting that mental files themselves are typed by the sort of ER relation they exploit. Different types of file — e.g. SELF-files, demonstrative files, recognitional files, deferential files, or what have you — exploit different types of contextual relation to the reference. On this view (i) the type of a file corresponds to its function or role: exploiting a given ER relation, and (ii) tokening a file of a given type

---

I saw John<sub>i</sub> the other day. The bastard<sub>i</sub> 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.

presupposes that the subject deploying the file is standing in the right relation to the reference. That presupposition corresponds to *a layer of content that comes in addition to the referential content of the thought in which the file is deployed*. It is that layer of content — cognitive content as opposed to referential content — that accounts for the subject's behaviour, and in terms of which the subject's rationality is to be assessed. For example, if the subject hallucinates a desirable object in front of him, the demonstrative file he deploys about 'it' is empty and fails to contribute to referential content (so the thought he entertains about the hallucinated object is neither true nor false) but the subject's attempt at grasping the object can be rationally accounted for in terms of the subject's presupposition that he is standing in front of a desirable object and gaining information from it in perception. Likewise, the subject's *saying* 'That is desirable' can be rationally accounted for in terms of the content of the subject's presupposition, even if the utterance itself fails to express a bona fide referential content.

Cognitive content is what is shared by two persons who both think e.g. 'My pants are on fire' (Kaplan 1989: 533). The referential contents of their respective thoughts are different (since A thinks that A's pants are on fire, while B thinks that B's pants are on fire), but they both deploy a SELF-file, and the sort of behaviour their thought causes is appropriate to the situation of someone whose pants are on fire. So, despite the difference in referential content, the cognitive content is the same.

Frege cases work in the other direction. The subject who looks at himself in the mirror and, not recognizing himself, thinks 'His pants are on fire', entertains a thought whose referential content is the same as that of the thought 'My pants are on fire', if he were to entertain it ; but the subject in that situation dissents from the latter (he does *not* accept 'My pants are on fire') while he assents to the former. So the cognitive content of the two thoughts, i.e. that which accounts for the subject's behaviour (including his assent/dissent behaviour), is different. In that situation the subject could say or think, without irrationality : 'His pants are on fire, but mine aren't.'

#### 4 Mental files as markers of identity

In the mental file framework, pioneered by Peter Strawson, identity statements do not involve descriptive senses, but instruct the hearer to *merge* the mental files respectively associated with the two singular terms flanking the identity sign (e.g. 'Cicero' and 'Tully'). As Strawson puts it, two 'segregated bundles or clusters of identifying knowledge' are 'brought together

and tied up into one for a given audience of an identity statement' (Strawson 1974 : 52; see also Lockwood 1971: 209). In general, when the subject realizes that two files concern the same object — whether or not this results from processing an identity statement — she should merge them into a single file.

In Frege cases, the subject has two distinct files about one and the same object, and that is what enables him or her to ascribe contrary predicates to that object without (internal) contradiction. The existence of distinct files in the subject's mind corresponds to his or her (mistaken) presuppositions that there are distinct objects out there. The subject's system of files thus expresses her ontology, and may embody two types of structural mistake. In cases of *confusion* one file in the subject's mind corresponds to distinct objects in reality. Confused files arguably fail to refer, since they track distinct objects at the same time (while wrongly presupposing that all information in the file is about the same object). In *Frege cases*, the situation is opposite : there is one object, but several files. Both types of case should be avoided, which means that the filing system is governed by the following norm :

***Strawson's constraint***

**One object, one file**

To comply with the norm, confused files have to be split up so as to match reality, while in Frege cases the files have to be merged. Mental files (qua particulars) thus serve as *internal markers of identity*. As Millikan puts it,

On Strawson's picture, the identity of a particular is represented in the mind by the identity of another particular. So long as you haven't made any mistakes, everything you know about your mother is attached to the same particular mental representation of your mother, the same token. Your understanding that all these facts are facts about the same woman consists in the representations of the logical predicates of each of these facts being attached to numerically the same "dot" in your mind or brain. Call this the "Strawson model" of how identity or sameness is thought. (Millikan 1997 : 500)

Among mental file theorists, some reject the Strawson model, or at least restrict its validity.<sup>12</sup> They distinguish two types of file. *Indexical files* are tied to particular ER relations, and track our contextually diverse relations to the reference. Strawson's 'one object, one file' constraint does not work for them, since it may be legitimate to take different perspectives simultaneously on the same object (by thinking of it through distinct files based on distinct ER relations). If I recognize the man I see in the mirror as myself, I entertain two representations of myself : as the man I see in the mirror, and as myself. When I think 'That man is myself' or 'I am that man', I simultaneously think of myself under these two distinct modes of presentation, which correspond to different relations to myself, while being aware that they are modes of presentation of the same person. No norm is thereby violated. Accordingly, when the subject realizes that the two files concern the same object, she does not have to merge them. What she has to do, rather, is to *link* them so as to let information flow freely between them (as if they were a single file).

On this alternative picture, Strawson's constraint only works for the other type of file : *encyclopedic files*. Encyclopedic files are not tied to any particular ER relations, they merely store any information we happen to possess about a given individual, whichever channel of information is or was exploited to gain that piece of knowledge. Rather than being associated with a particular ER relation, as indexical files are, encyclopedic files are only associated with a higher-order ER relation, namely the relation which holds whenever some first-order relation or other holds (Recanati 2012 : 73-74). Strawson's constraint works for encyclopedic files because there is no reason to entertain several encyclopedic files about the same individual : that could only betray the mistake of thinking that there are two distinct objects while there is only one. So when one realizes that two encyclopedic files concern the same object, one should indeed merge them into a single file.

The issue is far from settled, however (Recanati forthcoming). Even if we accept the distinction between indexical files and encyclopedic files, we can maintain that Strawson's constraint governs them all. For, once I recognize that the man I see in the mirror is myself, I can no longer think of 'him' as 'that man' ; I think of him as myself. The demonstrative way of thinking is *no longer available*, post-recognition. This supports the view that identity judgments cause the files to merge.

Be that as it may, Strawson's constraint is not needed to reach the conclusion that mental files are sameness markers. They have to be construed as sameness markers anyway,

---

<sup>12</sup> See Recanati (forthcoming) for an overview of that debate.

because *deploying the same mental file in thought several times represents the fact that the same individual is represented each time*. This is the widely debated phenomenon of coreference *de jure*, illustrated by unbound anaphora. To say that there is an anaphoric connection between two expressions none of which syntactically ‘binds’ the other is to say that the expressions are associated with the same mental file (token). This yields a strong form of coreference : internally represented coreference.

In cases of *de facto* coreference two token expressions happen to corefer, but the coreference relation is not internally represented via the identity of the mental file associated with both expressions ; rather, the speaker associates two distinct yet coreferential files with the two token expressions. In Frege cases, the subject refers to the same individual twice without knowing that the two acts of reference target the same individual. Because the subject does not know that the two files corefer, her ascribing contradictory properties to the same object (the referent of the two files) does not threaten her rationality. In cases of *de jure* coreference, however, the subject associates the same mental file with two token expressions, so the coreference relation is internally represented and cannot fail to obtain (assuming reference is achieved). 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).

The modal formulation (‘is bound to know’) is 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 well-known example :

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 while doubting that there is coreference or even wondering whether there is.<sup>13</sup> The coreference relation is entailed by the identity of the mental representations respectively associated with the two singular terms.<sup>14</sup>

---

<sup>13</sup> That is what we saw in the previous section, footnote 10, in connection with the ‘bastard’ example (‘I saw John<sub>i</sub> the other day ; the bastard<sub>i</sub> did not greet me’). Because of the anaphoric link, whoever understands the discourse knows that ‘John’ and ‘the bastard’ are bound to corefer.

<sup>14</sup> As Kit Fine says, ‘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).

## References

- Bach, K. (1987) *Thought and Reference*. Oxford : Clarendon Press.
- Bonardi, P. (forthcoming) Manifest Validity and Beyond. *Linguistics and Philosophy*.
- Castañeda, H.-N. (1999) *The Phenomeno-Logic of the I : Essays on Self-Consciousness* (edited by J. Hart and T. Kapitan). Bloomington : Indiana University Press.
- Chastain, C. (1975) Reference and Context. In K. Gunderson (ed.) *Language, Mind, and Knowledge*, Minneapolis : University of Minnesota Press, pp. 194-269.
- Crimmins, M. (1992) *Talk about Beliefs*. Cambridge, Mass. : MIT Press/Bradford Books.
- Crimmins, M. and Perry, J. (1989) The Prince and the Phone Booth. *Journal of Philosophy* 86: 685-711. Reprinted in Perry (1993), 249-78.
- Devitt, M. and Sterelny, K. (1999) *Language and Reality*, 2<sup>nd</sup> edition. Cambridge, Mass.: MIT Press.
- Dretske, F. (1988) *Explaining Behavior*. Cambridge, Mass. : MIT Press.
- Fine, K. (2007) *Semantic Relationism*. Oxford : Blackwell.
- Fodor, J. (1994) *The Elm and the Expert*. Cambridge, Mass.: MIT Press/Bradford Books.
- Fodor, J. (1998) *Concepts. Where Cognitive Science Went Wrong*. Oxford : Clarendon Press.
- Frege, G. (1984) *Collected Papers*, ed. B. McGuinness. Oxford : Blackwell.
- Garcia-Carpintero, M. and Torre, S. (eds.) (2016) *About Oneself*. Oxford : Oxford University Press.
- Gray, A. (2016) Minimal Descriptivism. *Review of Philosophy and Psychology* 7 : 343-64.
- Grice, P. (1969) Vacuous Names. In D. Davidson & J. Hintikka (eds.) *Words and Objections*, pp.118-45. Dordrecht: Reidel.
- Jackson, F. (1998) Reference and Description Revisited. *Philosophical Perspectives* 12: 201-18.
- Kamp, H. (1985) Context, Thought and Communication. *Proceedings of the Aristotelian Society* 85 : 239–261.
- Kaplan, D. (1989) Demonstratives. In J. Almog, J. Perry and H. Wettstein (eds.) *Themes from Kaplan*, pp. 481-563. New York: Oxford University Press.
- Kripke, S. (1979) A Puzzle About Belief. In A. Margalit (ed.) *Meaning and Use*, pp. 239-83. Dordrecht : Reidel.
- Kripke, S. (1980) *Naming and Necessity*. Oxford : Blackwell.
- Lewis, D. (1979). Attitudes *De Dicto* and *De Se*. *Philosophical Review* 88 : 513-43.

- Lewis, D. (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
- Lockwood, M. (1971) Identity and Reference. In M. Munitz (ed.) *Identity and Individuation*, pp. 199-211. New York : New York University Press.
- Mates, B. (1950) Synonymity. *University of California Publications in Philosophy* 25 : 201-26.
- Millikan, R. (1997) Images of Identity : In Search of Modes of Presentation. *Mind* 106 : 499-519.
- Pagin, P. (2013) The Cognitive Significance of Mental Files. *Disputatio* 5/36 : 133-45.
- Perry, J. (1979) The Problem of the Essential Indexical. *Noûs* 13: 3-21.
- Perry, J. (1993) *The Problem of the Essential Indexical and Other Essays*. New York: Oxford University Press.
- Perry, J. (2012) Thinking about the Self. In J. Liu & J. Perry (eds.) *Consciousness and the Self : New Essays*, pp. 76-100. Cambridge : Cambridge University Press.
- Pryor, J. (2016) Mental Graphs. *Review of Philosophy and Psychology* 7 : 309-41.
- Pylyshyn, Z. (2007) *Things and Places : How the Mind Connects to the World*. Cambridge, Mass. : MIT Press/Bradford Books.
- Recanati, F. (2012) *Mental Files*. Oxford : Oxford University Press.
- Recanati, F. (2013) Mental Files : Replies to My Critics. *Disputatio* 5/36 : 207-42.
- Recanati (2016) *Mental Files in Flux*. Oxford : Oxford University Press.
- Recanati (forthcoming) Do Mental Files Obey Strawson's Constraint? In C. Borgoni, D. Kindermann and A. Onofri (eds.) *The Fragmented Mind*. Oxford: Oxford University Press.
- Sainsbury, M. and Tye, M. (2012) *Seven Puzzles of Thought, and How to Solve Them: An Originalist Theory of Concepts*. Oxford: Oxford University Press.
- Searle, J. (1958) Proper Names. *Mind* 67 : 166-73.
- Sosa, E. (1995) Fregean Reference Defended. *Philosophical Issues* 6: 91-99.
- Stalnaker, R. (2008) *Our Knowledge of the Internal World*. Oxford : Clarendon Press.
- Strawson P. (1974) *Subject and Predicate in Logic and Grammar*. London: Methuen.
- Tałasiewicz, M. (2017) Mental files. Triggering Mechanisms, Metadata and "Discernibility of Identicals". *Studia Semiotyczne* 31: 13-35.
- Wittgenstein (1953) *Philosophical Investigations*, trans. E. Anscombe. Oxford : Blackwell.