

Mental Files and Identity

François Recanati
Institut Jean-Nicod

1. Introduction

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. For one and the same individual x , say Cicero, S may accept the claim that x was a philosopher if that claim is made in a certain way ('Cicero was a philosopher'), while rejecting the claim that x was a philosopher if it is made in a different way ('Tully was a philosopher'). Both 'Cicero was a philosopher' and 'Tully was a philosopher' say of the individual to whom both 'Cicero' and 'Tully' refer that he was a philosopher, so they make the same claim (true iff the individual in question was a philosopher), but the subject's acceptance or rejection of the claim depends upon the mode of presentation of the referent the two names share. If the referent is presented as Cicero, the claim is accepted, but if he is presented as Tully, it is not. That, of course, is possible only if S does not realize that Cicero *is* Tully. I assume that S has both the names 'Cicero' and 'Tully' in her repertoire, and that both names, as she uses them, refer to one and the same individual. The problem is that S herself does not know that. For her, there are two distinct individuals, and two distinct claims are made (one with respect to each of them).

To account for that sort of situation, Frege posited modes of presentation, or 'senses', in addition to the reference of linguistic expressions. And he appealed to this idea to account for the informativeness of identity statements such as 'Cicero is Tully'. At the level of reference, the statement is trivial, since an individual (the common referent of 'Cicero' and 'Tully') is said to be identical to that very individual — hardly a contingent matter. At the level of sense, however, the statement is informative precisely because the senses associated with 'Cicero' and 'Tully' are distinct. Sense determines reference, but does so only contingently. Because of that element of contingency, it is not guaranteed that the referents determined by two distinct senses (e.g. the sense of 'Cicero' and the sense of 'Tully') will be the same, and indeed a subject like S, unaware of certain contingent facts, takes them not to be the same. If the *senses* of the names were themselves the same (as in 'Cicero is Cicero'), the statement would be trivial and recognized as such by whoever understands it.

Now what *are* senses or modes of presentation ? Frege himself thought of them as essentially descriptive. The referent is presented as having certain properties or standing in certain relations to other entities. Since sense is supposed to determine reference, a unique object must have the relevant properties or stand in the relevant relations to other entities. So a sense can, in principle, be expressed by means of a *definite description* ‘the F’. The unique object which satisfies the descriptive condition (‘F’) is the referent. In cases such as ‘Cicero’ and ‘Tully’, the suggestion is that the subject, S, associates different descriptions with the two names.

There are good and well-known reasons for rejecting Frege’s descriptivist construal of senses, however. Senses are supposed to help us account for ‘Frege cases’. In Frege cases (e.g. ‘Hesperus’/‘Phosphorus’, ‘Cicero’/‘Tully’, etc.), the cognitive significance of two terms differs even though their reference is the same. (As a result, a rational subject may be led to ascribe contradictory properties to what is in fact the same object.) The problem with Frege’s descriptive take on senses is that, if accepted, it forces the theorist to posit reference-determining descriptions in the head of the subject *whenever* a Frege case is possible. Now there are three types of case in which contemporary philosophers of language and mind have been reluctant to do so :

1. *Reference through acquaintance*. 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). In many cases, we are actually 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).¹ 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.

Faced with such cases, the Fregean is likely to say that the reference-fixing description in the mind of the subject must be something like ‘what I am now seeing’ or ‘what I am now touching’. But this supposes, on the part of the subject, reflective

¹ As Campbell (2006 : 205) puts it, ‘Your visual system is managing to bind together information from a single thing, and you are consequently able to attend consciously to it, even though you have not managed to apply the right sortal concept to it’.

abilities the exercise of which is not required to suffer from identity confusions of the type which Frege cases illustrate. The subject need not reflect on her perceptual relation to objects in order to have thoughts about the objects she perceives ; nor does she have to reflect on her perceptual relation to objects to be in a position to think of the object in different ways, corresponding to the various ways in which she perceives it.

2. *Reference through communicative chains.* In cases such as the ‘Cicero’/’Tully’ case, the subject is able to describe the referent, but the descriptions he or she can provide do not fill the Fregean bill. First, the descriptions the subject can provide are often indefinite (‘a famous Roman orator’) rather than definite. That does not prevent the term(s) from referring. Second, when the subject is able to provide a definite description, the descriptive condition often fails to be satisfied by a unique object. Again, that does not prevent the term with which the description is associated from referring. Third, assuming the description the subject can provide is definite and uniquely satisfied, the satisfier need not be the referent of the term whose sense we are trying to characterize, as in Kripke’s Gödel/Schmidt case. This type of consideration has led a number of theorists, in the late sixties and early seventies, to argue in favour of an ‘externalist’ approach to reference determination. According to Geach, Kripke, Donnellan and others, what determines the reference of a name on a given use is not a description in the head of the users, but historical facts about that use and the communicative chain to which it belongs.

At this point the Fregean can make the following response : the description in the mind of the users in that sort of case is something like ‘the person called *Cicero*’, a description which (if the historical-chain picture is correct) is satisfied by whoever stands at the other end of the communicative chain which eventuates in the current use of the name. As Peter Geach pointed out, however, the *existence* of a communicative chain is sufficient to enable a name-user to successfully refer. The communicative chain does not have to be represented, any more than the perceptual relation to the referent has to be represented in order for the subject to successfully refer to an object he is acquainted with.

3. *Reference through indexicals.* Indexical modes of presentation are essentially perspectival and cannot be captured by means of objective, non-indexical descriptions.

As Castañeda and (following him) Perry forcefully pointed out, for any indexical α and non-indexical description ‘the F’, it is always possible for the subject to doubt, or to wonder, whether α is the F (Castañeda 1999, Perry 2000).

Reichenbach has suggested that an indexical is equivalent to a token-reflexive description (Reichenbach 1947 : §50). Thus a given token of ‘I’ presents its referent as the utterer of that token, a token of ‘now’ presents the time it refers to as including (or overlapping with) the time at which this token is uttered, etc. Insightful though it is, this move cannot support a descriptivist approach to indexical modes of presentation. What is needed to support such an approach is an objective, *non-indexical* description that provides the sense of the indexical. But for the token-reflexive description to count as non-indexical, the token in terms of which the referent is described must itself be described in objective/non-indexical terms, rather than referred to by means of a demonstrative like ‘this token’ (itself a variety of indexical). Now if the token is objectively described as, say, ‘the F-token’, the token-reflexive description will no longer be suitable for capturing the sense of the indexical. It is certainly possible for me to doubt that I am uttering the F-token, or to doubt that the F-token is being uttered now (or here), and that is sufficient to establish that the token-reflexive description (‘the utterer of the F-token’, ‘the time/place at which the F-token is uttered’ ...) does not provide the sense of the corresponding indexical (‘I’, ‘here’ or ‘now’). In any case, such token-reflexive descriptions can only be grasped by fairly sophisticated users of the language who, in addition to mastering the notion of ‘token’, are able to reflect upon the relations between token-representations and objects in the context in which these representations occur. Indexical thinking indeed exploits these relations, but in no way presupposes the ability to reflect on them.

To handle these three types of case, which are counterexamples to Frege’s descriptivist approach, we should make room for *non-descriptive* senses or modes of presentation. But what are such modes of presentation? What is it to think of an object non-descriptively? The theory of mental files is meant to answer that question. A non-descriptive mode of presentation is said to be a mental file. In Frege cases, the subject has two mental files for the same object (without realizing that there is a single object).

2. *Mental files*

2.1 *Mental files and ER relations*

There are different types of mental file, as we shall see in due course (section 3), but they all have the following properties. Mental files serve as individual or singular concepts. Like singular terms in the language, they refer, or are supposed to refer. What they refer to is not determined by properties which the subject takes the referent to have, but in externalist fashion, through relations (of the subject, or of the file itself construed as a mental particular) to various entities in the environment in which the file fulfills its function. The primary function of the file is to store information from these entities — information that is made available through the relations in question. The information (or misinformation) in the file therefore corresponds to the properties which the subject takes the referent to have, but these properties are not what determines the reference of the file.

The characteristic feature of the relations on which mental files are based, and which determine their reference, is that they are *epistemically rewarding* (hence my name for them : ER relations). They enable the subject to gain information from the objects to which he stands in these relations. In all the cases mentioned above as objecting to Frege's descriptivist approach, ER relations are involved. Relations of perceptual acquaintance are ER relations : they are the sort of relation to objects which makes the perceptual flow of information possible. Thus, by holding an object in my hand, I can get information about its weight ; by looking at it I can get information about its visual appearance. Perceptual files are, to use Perry's analogy, 'buffers' in which we store the information gained on the basis of these short-term relations. The relations of 'mediated acquaintance' established through communicative chains are also ER relations, which enable the subject (through communication) to gain information from the object at the other end of the communicative chain. The corresponding files are more enduring than perceptual buffers because the ER relation established through a communicative chain lasts longer than a transient perceptual relation.

The contextual relations to objects which indexical reference exploits are also ER relations, and they are typically short-lived, but not always. According to Perry (2002), the SELF file (which provides the sense of the indexical 'I') is based upon a special relation which every individual (permanently) bears to himself or herself, namely identity. In virtue of *being* a certain individual, I am in a position to gain information concerning that individual in all

sorts of ways in which I can gain information about no one else, e.g. through proprioception and kinaesthesia. The mental file SELF serves as repository for information gained in this way. In contrast, the files associated with the other indexicals ('here', 'now'...) are based on short-lived ER relations to the place we are in, or to the current time, which relations enable the subject to know (by using his senses) what is going on at the place or time in question. They are similar to perceptual buffers, which is to be expected given the link between indexicality and perception.

On the mental-file picture, what distinguishes descriptive from non-descriptive senses is the mechanism of reference determination. To use Kent Bach's useful terminology, reference determination is 'satisfactional' in the descriptive case, and 'relational' in the non-descriptive or *de re* case :

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 ; see also Bach 1986 : 188-9)

There is a variety of descriptivism which accommodates the relational nature of *de re* thought but 'internalizes' the relations by incorporating them into the content of the associated descriptions. On this view, the sense of a singular term always is that of a definite description, but in the allegedly 'non-descriptive' cases the descriptive condition F is relational. Thus, as we have seen, the descriptivist can say that in the perceptual case the mode of presentation is something like 'what I am seeing' or 'what I am touching'² ; and similarly for the other cases that raise *prima facie* difficulties for the descriptivist. But, as we have also seen, this view supposes reflective abilities the exercise of which is not actually required for having the relevant thoughts. The mental-file picture avoids this intellectualist pitfall. Mental files are based on relations to objects. Their function is to store information gained in virtue of standing in that relation to objects, and to represent them in thought. By deploying the file (or its 'address' or 'label') in thought, the subject can think about the object in virtue of standing in the relevant relation to it. But to entertain the thought, the subject does not have to reflect upon the relation in which she stands to the object.

² Or, in token-reflexive form : 'what is causing this visual/tactile experience'. (See e.g. Searle 1983 for a token-reflexive analysis of the content of perceptual experience.)

2.2 Linking files

Identity statements are informative, Frege says, whenever the senses of the terms on each side of the identity sign are distinct. Thus ‘A = B’ is informative, in virtue of the distinctness of the relevant senses, while ‘A = A’ is not, since (presumably) the same sense is exercised twice. In mental-file talk, this translates as follows : an identity statement ‘A = B’ is informative to the extent that the terms ‘A’ and ‘B’ are associated with distinct mental files. If the two terms are associated with the same file, the statement reduces to a (trivial) assertion of self-identity.

To say that there are two distinct mental files is to say that information in one file is insulated from information in the other file. Files are a matter of information clustering. Clustering takes place when all the information derives from the same source, through the same ER relation, and when it takes place, it licences the integration and inferential exploitation of the information in question. The role of the file is precisely to treat all the information as if it concerned one and the same object, from which it derives.³ But integration and exploitation of information is blocked if the relevant information is distributed in distinct files, for then, there is no presumption that all the information derives from the same object. So, even if I know that Cicero is bald, and that Tully is well-read, I cannot conclude that some bald man is well-read, despite the fact that Cicero is Tully : the information ‘is bald’ is in the Cicero file, while the information ‘is well-read’ is in the Tully file. Informational integration and inferential exploitation of information only takes place within files, on this picture.

There is, however, an operation on files whose role is precisely to overcome that architectural limitation, by licensing the integration/exploitation of information distributed in distinct files. That operation, following Perry, I call *linking*. When two files are linked, information can flow freely from one file to the other, so informational integration/exploitation becomes possible (despite the constraint that it can only occur within

³ ‘Updating one’s files involves being disposed to collect information as if there is some one individual that one’s file F has always been about. One’s screening and pruning dispositions are responsive to this purported fact’ (Lawlor 2001 : 88). ‘One reason for not allowing an individual concept [= a file] to change its referent is that the referent fixes a condition for the coherence of information within an individual concept : if ‘is F’ belongs in belief mode to a given individual concept, then ‘is not F’ should not. The constraints on updating would not obtain if an individual concept might shift its referent’ (Sainsbury 2005 : 232). For a detailed argument to the same effect, see Millikan 1997 : 504-506, 2000 : 141-144.

files). Thus if I learn that Cicero is Tully, this allows me to put together the pieces of information in the two files, and to infer that some bald man is well-read.

From a cognitive point of view, linking is a quite fundamental operation. It is involved, for example, in the phenomenon of recognition (which involves linking a perceptual file and a file based on memory). It is that operation which I think accounts for the cognitive effect of accepting an identity statement. To accept the identity 'A = B' is to link the two files corresponding to the terms on each side of the equals sign. It would be incoherent to accept the identity 'Cicero = Tully', and not let the information in the respective files get together and breed.

2.3 Linking vs merging

Strawson was the first to claim that identity judgments should be understood in terms of their effects on the management of files in the mind of the thinker (Strawson 1974 : 51-56). Two 'segregated bundles or clusters of identifying knowledge', he says, are 'brought together and tied up into one for a given audience of an identity statement' (p. 52). But the operation on files which, according to Strawson, results from accepting an identity statement is the 'merge' operation through which the two files become one. ('Merge' is Millikan's term, not Strawson's.) As several authors noticed, however, the 'merge' model is not adequate to describe the cognitive effects of identity judgments.

Two linked files may end up being merged, after some time (especially as new information accumulates), but there are all sorts of reasons also for not automatically merging two files that are linked (Lawlor 2001 : 62-65 and 92-93). For example, it would be very risky to merge two files on the basis of an identity judgment that one may accept with less than 100% subjective probability (Millikan 1997 : 508). Linking is less risky, as it can easily be undone. So merge is an option for dealing with an identity, but it should not be automatic.⁴ Second and most importantly, the 'merge' model is incompatible with the mode of presentation idea we are trying to cash out (Millikan 2000 : 147-149). It is of the essence of modes of presentation that there can be a multiplicity of modes of presentation for the same object. On the picture I have presented, mental files qua nondescriptive modes of presentation correspond to various relations in which the subject stands to objects, and there is no doubt whatsoever that a subject can and typically does stand in several relations simultaneously to

⁴ 'If the identification [A = B] is tentative, the notions [= files] may retain their identity ; if not, they may merge and become one' (Perry 2002 : 196).

the objects in his or her environment. Nor is this situation contrary to some normative ideal, as if the coexistence of several files for a single object was a defect to be avoided whenever possible. Imagine that I see a certain man cutting his grass and recognize him as Noam Chomsky. (Or imagine I learn he is Noam Chomsky, through an identity statement which I accept.) My perceptual file and my Chomsky file get linked, but there is no reason why either should disappear. Perry describes the perceptual buffer as being ‘absorbed’ into the more permanent file in such cases, but I think the buffer should only disappear when the ER relation on which it is based no longer holds. Taking seriously the idea that mental files are modes of presentation based upon contextual relations to objects demands that we accept the existence of a multiplicity of files for the same object even when the files are linked and the subject is aware that they stand for a single object.

Strawson’s idea that two linked files should be merged makes sense in the context of his own enquiry, however. He was concerned with a very specific type of file, associated with proper names (his topic in the relevant passage of *Subject and Predicate in Logic and Grammar*). I call that type of file an ‘encyclopedia entry’ (Recanati 1993, 2010 ; see §3.3 below). Encyclopedia entries do obey the norm that there should be exactly one per object of interest. In *Reference and Reflexivity*, Perry describes files as ‘little cards in the mind on which we jot down information about people, things and places. My picture is *a card for each person, place, or thing*’ (Perry 2001b : 54 ; emphasis mine). This description fits encyclopedia entries well, though I think it is a mistake to apply it to files in general. I shall say more about encyclopedia entries, and the norm that governs them, in section 3. For my present purposes, the important point is that encyclopedia entries are only one particular type of file, and that the norm in question only applies to that type of file. It does not apply to files in general, hence there is no reason to accept that, in general, linking does or should give rise to a merging of files.

2.4 Presumptions of identity within files

Besides judgments of identity, the cognitive effect of which is to link two files, there are also *presumptions* of identity, whose status is quite different. While linking only operates on distinct files, presumptions of identity are operative within a single file. As I wrote above, to put various pieces of information in the same file means that they are supposed to derive from a single source (i.e. through some ER relation to a given object). Pieces of information in the same file can thus be inferentially integrated and exploited as if they concerned the same

object (whether or not they actually do). This Campbell describes as ‘trading on identity’ (Campbell 1987, 1994, 2002 ; see also Millikan 1997, Fine 2007).

It is tempting to regard presumptions of identity as nothing but *implicit* judgments of identity. On this view the difference between argument A and argument B below is that, in argument A, the judgment of identity is explicit, while it remains implicit in argument B (which is therefore enthymematic). The reason why it can remain implicit in B is that the identity is obvious and trivial, so it ‘goes without saying’ and can be suppressed, in contrast to what happens in argument A.

<i>Argument A</i>	<i>Argument B</i>
Cicero is bald	Cicero is bald
Tully is well-read	Cicero is well-read
Cicero = Tully	[implicit premise : Cicero = Cicero]
Someone is bald and well-read	Someone is bald and well-read

The suppressed premise in argument B is meant to ensure that the two occurrences of ‘Cicero’ in the explicit premises of the argument actually corefer (if they did not corefer, the argument would be invalid, indeed). As Campbell and many others have shown, however, this view of argument B as enthymematic and resting on a suppressed premise is indefensible. In general, the attempt to reduce presumptions of identity to implicit identity judgments launches an infinite regress :

If this view were correct, we would also need to make sure that the uses of [‘Cicero’] in the suppressed premise are linked with the uses of [‘Cicero’] in the explicit premises, and we would need further suppressed premises to secure these connections. The problem recurs, and we are embarked on a regress. (Campbell 1994 : 75)

This means that we cannot regard arguments like the one under consideration as enthymematic, needing but a further (object-language) sentence to be made completely valid ; there is no evading unthinking reliance on sameness of reference. (Sainsbury 2002 : 135)

According to the [suggestion]... what it is to think that the individual Cicero is a Roman and then to have the coordinated thought that he is an orator is to think the additional thought that the one individual is the same as the other. But if the new thought is to have the desired effect, then it must be supposed that the individuals in the new thought are represented as the same as the respective individuals in the original thoughts ; and so the account is circular. (Fine 2007 : 68)

To insist that the subject must make an explicit identity judgment before she can recognize that two thoughts are about the same thing would be to invite a vicious regress—for even the simplest inference from ‘P’ to ‘P’ would then require infinitely many explicit identity judgments to establish the co-reference of premise and conclusion. The moral here is much the same as the one Lewis Carroll drew in the case of modus ponens: we must have some basic way of taking two thoughts to be co-referential which does not require an explicit identity judgment. (Schroeter 2008 : 115n)

I conclude that we need both the identity presumptions (or, according to the now standard terminology, ‘coreference *de jure*’) and identity judgments (‘coreference *de facto*’).⁵ In the mental-file framework, they correspond respectively to the clustering of information into files, and the linking of (distinct) files.

⁵ Fine (2007) distinguishes ‘thinking of something as the same’ (*de jure* coreference, in the standard terminology, or, in Fine’s own terminology, ‘strict coreference’) and ‘thinking of something as *being* the same’. ‘A good test of when an object is represented as the same 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). However, as we shall see in section 4, *de jure* coreference as characterized through this test is an overly broad notion which covers different types of case, including *some* cases in which distinct files are involved.

3. *The hierarchy of files*

3.1 *Proto-files and the Generality Constraint*

I have alluded to the existence of several types of file. So far, the files I have talked about (with the exception of ‘encyclopedia entries’, on which more below) are very closely tied to specific ER relations on which they are based. The file exists only as long as the relation (hence the possibility of gaining information about the object by exploiting the relation) exists, and for that reason the life expectancy of many files is rather short : they are *temporary* files. So, as long as I am in the right type of perceptual contact with the grass-cutting man, I can think of him demonstratively (‘that man’). When I am no longer in a position to perceive him or to focus my attention on him, I can no longer think of him under the demonstrative mode of presentation since the latter involves the activation of a mental file which depends upon the existence of the right type of perceptual relation. When the relation is broken, the temporary file based on it disappears. (The information in the file is not lost, of course, but transferred into other files, e.g. the Chomsky file to which the demonstrative buffer is linked.)

At this point we need to introduce a distinction between two types of file based on ER relations. ‘Proto-files’, as I am going to call them, can *only* host information gained in virtue of the ER relation to the referent. For example, the proto-file SELF* can only host information gained ‘from inside’, in the first person way ; the demonstrative proto-file THAT MAN* can only host information gained by perceptually attending to the object. I call these files ‘proto-files’ (and mark them with an asterisk) rather than files *simpliciter* because I want files (properly speaking) to serve as individual *concepts*, i.e. thought constituents ; and I take proto-files to lack a distinguishing characteristic of concepts.

Concepts, in general, satisfy or ought to satisfy what Evans calls the Generality Constraint. Evans says that a subject in possession of a predicative concept *F* should be able to entertain thoughts in which that concept is applied to any object of which the subject has an individual concept ; similarly,

If a subject can be credited with the thought that *a* is *F*, then he must have the conceptual resources for entertaining the thought that *a* is *G*, for every property of being *G* of which he has a conception. This is the condition that I call ‘The Generality Constraint’. (Evans 1982 : 104)

In the mental-file framework, predication is cashed out as follows : the file is what stands for the object of which something is predicated, and the predicate's location within the file means that it is taken to apply the object in question. Translated into mental-file talk, the Generality Constraint says that a file should be hospitable to any predicative concept in the subject's possession. Clearly, that is a constraint which proto-files do not satisfy. Take the proto-file SELF* : it can only host information gained from inside, through e.g. proprioception or introspection. Now there is much information about myself that I cannot gain in this way. My date of birth is something I learn through communication, in the same way in which I learn my parents's birthdates. In virtue of the Generality Constraint, it should be possible for that information to go into my SELF file, and that is the crucial difference between the SELF file and the (nonconceptual) proto-file SELF* from which it originates.

3.2 Conceptual files

In contrast to proto-files, which are based on some ER-relation and can only host information derived through that relation, a (conceptual) file based on a certain ER relation contains two sorts of information : information gained in the special way that goes with that relation (first-person information, in the case of the SELF file), and information not gained in this way but *concerning the same individual as information gained in that way*. Information about my birthdate is a case in point : I gain that information in a third-person way, through communication (as I might come to know anybody's birthdate), but I take that piece of information to concern the same person about whom I also have direct first-person information, i.e. myself ; so that information, too, goes into the SELF file. I am therefore able to exercise my SELF concept in thinking 'I was born in 1952'.

It is because of that dual aspect of the SELF concept *qua* satisfier of the Generality Constraint that there are two types of 'I'-thoughts : those that are, and those that are not, immune to error through misidentification. When some information is gained from inside, that is, in virtue of the ER relation on which the SELF file is based, that information can only be about the subject : the way the information is gained determines which object it concerns (or, equivalently, in which file it goes). As a result, as Evans puts it, 'there just does not appear to be a gap between the subject's having information (or appearing to have information), in the appropriate way, that the property of being *F* is instantiated, and his having information (or appearing to have information) that *he* is *F*' (Evans 1982 : 221). But when some information about ourselves is gained from outside, it goes into the SELF file only in virtue of a judgment

of identity. The thought ‘I was born in 1952’ can thus be seen as the product of two thoughts : the thought that a certain person x , namely the person I hear about in a given episode of communication, was born in 1952, and the thought that I am x . The thought that I was born in 1952 thus turns out to be ‘identification-dependent’, in Evans terminology.

Despite the fact that they can host information not derived through the ER relation on which the file is based, the file is still based on that relation. What this means is that *the file exists only as long as the relation exists* (and with it the special way of gaining information about the referent through the relation). Files should therefore be seen as an expansion of proto-files, including the proto-files themselves as their nucleus.⁶ Linking is what necessitates the expansion of the original proto-files. Linking makes information flow between files, but that only makes sense if the files can accept ‘alien information’ (information from other files, that is, information that is not derived through the ER relation on which the file is based). So linking and expansion of proto-files are best construed as two sides of the same coin.

3.3 Encyclopedia entries

The distinction between proto-files and conceptual files based on the Generality Constraint was drawn in *Direct Reference* (Recanati 1993, chapter 7). Another important distinction, also made in chapter 7 of *Direct Reference*, is between two types of conceptual files. The files I have talked about so far are what we might call ‘first-order’ files. They are based on specific ER relations. But we must allow room for more abstract files, based on the relation to an object that holds when we have first-order files about it, that is, when we stand in various ER relations to it which enable us to gain information from it in various ways. In Recanati (2010) I characterize such files as follows :

Not all files are based on specific contextual relations enabling us to gain information about the referent in particular ways. Some files (the *indexical* files) are based on specific contextual relations, such as one’s relation of identity to oneself or the relation to what we hold in our hand, but others (the *encyclopedic* files) are based on a more

⁶ Peacocke notes that when the subject falls prey to a perceptual illusion of which he is aware, the content of the illusion ought not to figure in the subject’s conceptual file about himself, because, at the level of judgment, ‘the subject rejects the content of his more primitive, pre-judgemental phenomenology’ (Peacocke forthcoming). This might be taken to argue against the inclusion of the primitive proto-file within the conceptual file. I acknowledge the difficulty, but cannot discuss the issue here.

general-purpose tracking relation. Thus my file about Mont Blanc contains all the information I can get about the mountain, *however it is gained*. It is not tied to a particular way of gaining information, nor to a specific ER relation. An encyclopedic file may exploit a number of ER relations to the reference of the file, in an opportunistic manner, instead of being based on a single one. Any relation will do, provided it preserves the link to the object. In this case, what determines the reference of the file is the overarching tracking relation : the relation between the file and the object it has been created to track (however it *is* tracked). Not being based on a specific ER relation, an encyclopedia entry is not short-lived, as the other type of file typically is. It survives when our contextual relation to the reference changes. (Recanati 2010 : 157-158)

This might suggest that encyclopedic files are not based on ER relations ; however, that is not what I want to say. Rather, I distinguish between specific ER relations and the higher-order ER relation on which encyclopedia entries are based, namely

$$\lambda x \lambda y [(\exists R) (Rx,y)]$$

where ‘R’ ranges over ER-relations. A subject (or a mental file in the subject’s mind) x stands in that relation to an object y just in case there is/are some first-order ER relation(s) in which x stands to y . A file based on the higher-order relation hosts any information derived in virtue of that relation, that is, ultimately, any information derived in virtue of any of the first-order ER relations. Such files correspond to what Perry calls ‘detached’ files.⁷

For encyclopedia entries, the Strawsonian constraint ‘one object, one file’ holds : since the file abstracts from the specific ER relations, there is no point in entertaining distinct encyclopedia entries about the same object. All such files would be based on the same relation to the object (the higher-order relation) so their multiplicity could only reflect the mistake of thinking that there are two objects where there is one.

⁷ ‘Think of the architecture of our beliefs as a three-story building. At the top level are detached files.... At the bottom level are perceptions and perceptual buffers. Buffers are new notions associated with the perceptions and used to temporarily store ideas we gain from the perceptions until we can identify the individual, or form a permanent detached notion for him, or forget about him. The middle level is full of informational wiring. Sockets dangle down from above, and plugs stick up from below’ (Perry 2001a : 120-21).

I speak of a ‘hierarchy’ of files for two reasons. First, files at each of the three levels I have described presuppose files at the previous level. Proto-files are the most basic ; conceptual files are generated from them (through linking-*cum*-expansion). Among conceptual files, first-order files are more basic, since higher-order files — encyclopedia entries — presuppose them. Second, files can be ordered in terms of how closely tied they are to specific ER relations. Proto-files are very closely tied to specific ER relations since they can only host information derived on the basis of these relations. This constraint is relaxed in conceptual files. Still, conceptual files remain closely tied to ER relations in the sense that their existence (like that of proto-files) is conditional upon the existence of specific ER relations to the referent. This second constraint is relaxed in encyclopedia entries, since they do not depend upon specific ER relations for their very existence, but only on there being some ER relation or other to the referent. Still, all types of file, including encyclopedia entries, are based upon ER relations to the object they are about, and their reference depends upon the ER relations on which they are based.

4. Objections and Responses

4.1 The circularity objection

Two of my PhD students, Michael Murez and Gregory Bochner, have argued that the mental-file account of *de jure* coreference is circular.⁸ Let me start by quoting a passage from Bochner’s paper :

Many advocates of mental files, while presenting their view, themselves acknowledge—as if this were compatible with what they claim—that a mental file is created when an object is *taken to be* one by the subject.⁹ (...) But to claim that a file

⁸ Murez 2009 ; Bochner 2010. The circularity worry is already expressed in this passage from Lawlor’s dissertation : ‘In the context of building from the ground up an account of what constitutes coreferential thinking... [the theorist] owe[s] an account of what makes information belong to a single file. And [she] cannot provide this account in terms of a thinker’s capacity for coreferential thinking. That would be viciously circular’ (Lawlor 2001 : 80).

⁹ At this point Bochner quotes Forbes, who writes : ‘When we receive what we take to be *de re* information which we have an interest in retaining, our operating system may create a locus, or dossier, where such information is held; and any further information which we take to be about the same object can be filed along with the information about it we already

is created for what is (*perhaps mistakenly*) taken to be an object is just to acknowledge that co-reference *de jure* rests on something akin to a prior judgement, [and this... suggests that *de facto* co-reference is primitive]. If you already need to think of the object *in order to* determine that it is a single object deserving a single location in your syntax, then this means that you must be able think of the object *prior* to the attribution of a vehicle or mental file. And, presumably, if some identity mistake is made in this early process of syntactic assignment—if, for instance, two different vehicles are created for a unique object taken to be two distinct objects—it will be *that* early mistake that will explain cognitive significance, not the fact that there are two vehicles. (...) All of this is incompatible with the idea... that it is differences in syntax that determine differences in cognitive significance, and, instead, squarely supports the opposite view that it is differences in cognitive significance that determine differences in syntax.

As I understand it, the objection targets both the mental-file account of *de jure* coreference, and the mental-file account of identity judgements. According to the mental-file account as I have presented it, there is *de jure* coreference when two pieces of information occur in the same file and are ‘presumed’ to be about the same object; while identity judgements have the effect of linking two distinct files, thereby establishing *de facto* coreference between the pieces of information in the respective files (e.g. ‘is bald’ and ‘is well-read’ in our earlier example). But if *de jure* coreference is a matter of belonging to a single file, and if, in turn, belonging to a single file is a matter of ‘being taken to be about the same object’, then (i) *de jure* coreference rests on judgments of identity and therefore reduces to *de facto* coreference, and (ii) the account of *de facto* coreference (judgments of identity) in terms of operation on files leads us into a regress, since the files themselves, presuppose identity judgments : what goes into the file is whatever information we gain concerning *the same object as* information already in the file.

In response, let me start by noting that the central idea that various pieces of information cluster into a single file when they are ‘taken to concern the same object’ can be understood in a way that does *not* presuppose a prior identity judgment. It may be entirely a

possess. [...] The role of a name is to identify a file for a particular object—as I shall put it, we use names to “label” dossiers. In sum, then, on coming across a new name, one which is taken to stand for some particular individual, the system creates a dossier labeled with that name and puts those classified conditions into it which are associated with the name.’ (Forbes, 1990: 538; Bochner’s emphasis.)

matter of sub-personal binding of information. Thus in the case of proto-files at least it is the cognitive system, not the subject, that takes the pieces of information to concern the same object and cluster them within a file. The subject does not judge that the pieces of information concern the same object. Identity is presupposed, it is built into the way the information is (subpersonally) packaged. So this is very different from an identity judgment, and coreference *de jure* does not reduce to coreference *de facto*.

As Murez makes clear, however, the difficulty comes from the introduction of *conceptual* files (the sort of files which are thought constituents and which are exercised when we make judgments). Something goes into a (conceptual) file if it is *judged* to concern the same individual as information in the file. The circularity objection therefore holds against what Murez calls the ‘sophisticated theory’ of mental files, i.e. the theory that goes beyond proto-files and makes room for conceptual files (which Murez refers to as ‘relation-independent files’). Judgments of identity are accounted for in terms of a certain operation on such files (linking), yet the files in question, because they are conceptual, constitutively depend upon certain identity judgments : what goes into the file is whatever information we gain concerning (what we take to be) *the same object as* information already in the file. But if conceptual files themselves depend upon identity judgments (in order to be fed the ‘alien information’ they need to qualify as conceptual files), then we *cannot* analyse identity judgments in general in terms of a linking operation on files, as I have done, without launching a regress.

4.2 Proto-linking

To avoid the regress, we must take advantage of the hierarchy of files and the distinction of levels it is based on. Proto-files, as Murez acknowledges, are characterized in terms of the ER relation which determines which information is allowed into the file’s content, independently of any identity judgment. So what we must do is introduce a clear distinction between linking as it operates on *these* files (the proto-files), thereby making it possible for them to achieve the status of conceptual files, and linking as it operates on the conceptual files that result from the prior linking operation. The prior operation we can refer to as ‘proto-linking’ (it corresponds to what, earlier, I called ‘linking-cum-expansion’). This distinction provides a way to avoid the regress. Identity judgments are accounted for in terms of an operation on (conceptual) files, namely linking. The very notion of a conceptual file itself presupposes some linking operation, but there is no circularity because the linking operation which the

notion of a conceptual file presupposes — proto-linking — is not quite the linking operation which accounts for identity judgments. It operates on proto-files, while linking operates on conceptual files ; and it expands the proto-file beyond (what becomes) the nucleus, while linking does not affect the structure of the conceptual files it operates on.

Of course, there is a sense in which it is the same operation of linking in both cases, namely the operation that connects two files and makes it possible for information to flow freely between them ; but in one case it operates on proto-files, forcing alien information into them and expanding them accordingly, while in the other case it operates on already constituted conceptual files with a dual structure (nucleus + periphery) which it exploits but does not affect or modify. That difference — between the two cases is sufficient to meet the objection. The files on which proto-linking operates are proto-files, whose proper functioning does *not* rest on identity judgments ; hence no circularity is involved in analysing identity judgments as establishing a link between two conceptual files, which themselves are analysed as resulting from a (proto-)linking operation on proto-files.¹⁰

4.3 *The transitivity objection*

In a couple of recent papers (Pinillos 2009, forthcoming), Angel Pinillos argues that ‘third object’ accounts of *de jure* coreference (of which the mental-file account is an instance) are bound to fail. According to such accounts, Pinillos says, two terms are coreferential *de jure* if and only if they are associated with a single entity (e.g. a single mental file) which constitutes or determines their shared cognitive significance. (That entity is the ‘third object’.) Pinillos’s alleged knock-down objection to such accounts is that being coreferential *de jure* is not a

¹⁰ Earlier, I mentioned recognition as a rather fundamental cognitive phenomenon involving linking. Presumably, recognition exists in non-conceptual creatures, and if it does, it is an instance of proto-linking. This suggests that proto-linking by itself is not sufficient to turn proto-files into conceptual files ; for if it were, it would be incoherent to assume that non-conceptual creatures can have recognitional capacities (since proto-linking would automatically endow them with concepts). There are, indeed, several reasons for decoupling proto-linking from conceptualisation. First, the limited intake of alien information that comes with proto-linking is likely to be only a first step towards satisfying the full-blooded Generality Constraint. Second, more constraints than the Generality Constraint presumably have to be satisfied for conceptual thought to emerge. Instead of assuming that proto-linking automatically converts proto-files into conceptual files, therefore, we should enrich the hierarchy of files with an extra level : between the proto-files and the conceptual files, we should posit an intermediate category, namely the expanded proto-files which result from proto-linking. Such files remain non-conceptual yet they are hospitable to alien information (to some extent at least) and go some way toward satisfying the Generality constraint.

transitive relation. It is possible for A and B, and for B and C, to be coreferential *de jure*, even though A and C are not. But if the relation of *de jure* coreference rested on the identity of the mental files respectively associated with each of the terms, it should be transitive, since identity is a transitive relation. Pinillos takes his argument to support Fine's 'relationist' approach to *de jure* coreference (Fine 2007).

Pinillos gives examples like the following to show that *de jure* coreference is not a transitive relation :

- (0) We were debating whether to investigate both Hesperus₁ and Phosphorus₂; but when we got evidence of their true identity, we immediately sent probes there_{1,2}.
- (1) As a matter of fact, my neighbor John₁ is Professor Smith₂, you will get to meet (the real) John Smith_{1,2} tonight.
- (2) Hesperus₁ is Phosphorus₂ after all, so Hesperus-slash-Phosphorus_{1,2} must be a very rich planet.

He argues that in each example, there are three terms A, B and C such that A and B are coreferential *de jure*, B and C also are coreferential *de jure*, yet A and C are only coreferential *de facto*. In (0) and (2), 'Hesperus' and 'Phosphorus' are coreferential *de facto* (as they feature in the informative identity judgment 'Hesperus is Phosphorus'), yet both 'Hesperus' and 'Phosphorus' are *de jure* coreferential with 'Hesperus/Phosphorus' in (2) or with the anaphoric 'there' in (0). As Pinillos puts it,

'There' and 'Hesperus' are *de jure* coreferential because... one who understands the use of (0) must know that if 'there' refers at all it must corefer with 'Hesperus'. Hearers who understand the use of (0) know that those occurrences can't refer to different objects. Similarly, hearers who understand the use of (0) must also know that if 'there' refers at all, it must corefer with 'Phosphorus'. Hearers must know that they can't refer to different things. Hence, (...) 'there' is *de jure* coreferential with both 'Hesperus' and 'Phosphorus'. (Pinillos 2009)

The criterion for *de jure* coreference which Pinillos appeals to here is the knowledge criterion: when two terms are *de jure* coreferential, one cannot understand the utterance without knowing that the two terms corefer (if they refer at all).

So example (0) shows that *de jure* coreference is not a transitive relation. Pinillos

concludes that third objects account fail. According to such accounts, he says,

Occurrences A and B in a discourse are *de jure* coreferential because they stand in a certain relation R to a single object X (e.g. a single mental file).

This cannot be right, Pinillos points out, for if it were — if *de jure* coreference was a matter of identity (identity of mental file, say) — it would be a transitive relation ; but example (0) establishes that it is not.

4.4. Reply to Pinillos

I agree with Pinillos that example (0) shows that *de jure* coreference, as characterized through the knowledge criterion, is not a transitive relation ; but I deny that this argues against the mental-file account (or third-object accounts in general). It does argue against a strong version of such accounts, namely :

Occurrences A and B in a discourse are *de jure* coreferential (i.e. pass the knowledge test) *just in case* they stand in a certain relation R to a single object X (e.g. a single mental file).

But it does not argue against a weaker version of the theory, according to which being associated with a single mental file (or any relevant third object) is sufficient, though not necessary, for *de jure* coreference :

Occurrences A and B in a discourse are *de jure* coreferential if they stand in a certain relation R to a single object X (e.g. a single mental file).

This weaker version leaves open the possibility that two occurrences A and B might be *de jure* coreferential (i.e. pass the knowledge test) for some other reason than their being associated with the same mental file.

The knowledge test characterizes *de jure* coreference in terms of a priori knowledge of coreference. Now, if two terms have the same sense, understanding the terms (knowing their sense) entails knowing that they corefer ; that follows from the constraint that sense determines reference. So identity of sense entails *de jure* coreference. But why should the entailment be bidirectional ? There may be other sources of a priori knowledge of coreference

than sense-sharing. Indeed, I will argue, that is exactly what is going on in Pinillos's examples : a priori knowledge of coreference is secured, but the source of such knowledge is not the identity of the associated mental files (or the identity of sense, more generally).

In each example we find that two terms (say, 'Hesperus' and 'Phosphorus') are associated with distinct files — so they are only *de facto* coreferential — but there is also a third file (say the 'Venus' file or, better, the 'Hesperus/Phosphorus' file) which is created when one learns that the two terms actually corefer. That third file is what the 'merge' model posits : it says that, upon understanding and accepting an identity, one feeds all the information from the two initial files into a third file, and suppresses the initial files.¹¹ On the weaker 'link' model, one does not (automatically) suppress the initial files, but that does not prevent one from opening a file for the unique object which is the referent of the two initial files. It is the function of slash-terms such as 'Hesperus/Phosphorus' to be associated with such files. On the 'link' model, it is possible to construe such files as including each of the initial files (now connected though linking) as sub-files. Be that as it may, given the way the inclusive file is introduced and its role, it is a priori that it corefers with each of the initial files. So it is a priori that 'Hesperus' and 'Hesperus/Phosphorus' corefer (if they refer at all). Likewise it is a priori that 'Phosphorus' and 'Hesperus/Phosphorus' corefer (if they refer at all). It follows that both 'Hesperus' and 'Phosphorus' are *de jure* coreferential with 'Hesperus/Phosphorus'. Yet these terms are *not* associated with the same file : 'Hesperus' and 'Phosphorus' are associated respectively with what I called the 'initial files', while 'Hesperus/Phosphorus' is associated with what I called the 'inclusive file'.¹² In this case, therefore, we have an instance of *de jure* coreference that is not accounted for in terms of a shared file, but in terms of a relation other than identity between two distinct files : the

¹¹ 'On receipt of an identity-statement invoking two... clusters, the two appropriate cards are withdrawn and a new card is prepared, bearing both the names of which one heads one of the original cards and one the other, and incorporating the sum of the information contained in the original cards ; the single new card is returned to stock and the original cards are thrown away' (Strawson 1974 : 56).

¹² As Fine notes, 'it is not that the merged file represents the individual as the same as the earlier files, since that would require that the earlier files represent the individual as the same. Rather, the new file, *if* I choose to create it, will represent the individual as *being* the same as the earlier files' (Fine 2007 : 69). In this passage in which he talks about the mental-file account, Fine seems to acknowledge that 'Hesperus' and 'Hesperus/Phosphorus' are not 'strictly coreferential' (even though they pass the 'knowledge test' he himself uses — see footnote 5). Strict coreference, thus understood (i.e. narrowly), corresponds to the idea that the two terms are associated with a single file. (Fine argues against accounts of strict coreference in terms of either mental files or nondescriptive senses, but his arguments do not convince me, and, as far as I can tell, his criticism does not apply to the present account.)

relation — whatever it is exactly — that holds between the initial files and the inclusive file. Such cases can be accounted for within the mental-file account, so they do not argue against it ; they only argue against an implausibly strong version resting on the (unargued) premise that only the identity of associated files can be responsible for a priori knowledge of coreference.¹³

¹³ This paper was my Gareth Evans Memorial Lecture (Oxford, 25 January 2011) ; I am happy to dedicate it to Kevin for his birthday. The research leading to the lecture has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement n° FP7-238128 and ERC grant agreement n° 229441-CCC.

References

- Bach, K. (1986) Thought and Object : *De Re* Representations and Relations. In M. Brand & R.M. Harnish (eds), *The Representation of Knowledge and Belief*, pp. 187-218. Tucson : The University of Arizona Press.
- Bach, K. (1987) *Thought and Reference*. Oxford : Clarendon Press.
- Bochner, G. (2010) Cognitive Significance and Non-Descriptive Senses. *Ms*.
- Campbell, J. (1987) Is Sense Transparent ? *Proceedings of the Aristotelian Society* 88 : 273-92.
- Campbell, J. (1994), *Past, Space and Self*. Cambridge, Mass. : MIT Press.
- Campbell, J. (2002) *Reference and Consciousness*. Oxford : Oxford University Press.
- Campbell, J. (2006) Sortals and the Binding Problem. In F. McBride (ed) *Identity and Modality*, Oxford : Clarendon Press, 203-18.
- 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.
- Donnellan, K. (1970) Proper Names and Identifying Descriptions. *Synthese* 21 : 335-58.
- Dretske, F. (1988) *Explaining Behavior*. Cambridge, Mass. : MIT Press.
- Edelberg, W. (1995) A Perspectivalist Semantics for the Attitudes. *Noûs* 29 : 316-42.
- Evans, G. (1982) *The Varieties of Reference* (ed. by J. McDowell). Oxford : Clarendon Press.
- Fine, K. (2007) *Semantic Relationism*. Oxford : Blackwell.
- Forbes, G. (1990) The Indispensability of *Sinn*. *Philosophical Review* 99: 535-563.
- Frege, G. (1984) *Collected Papers on Mathematics, Logic, and Philosophy*. Oxford : Blackwell.
- Geach, P. (1972) *Logic Matters*. Oxford : Blackwell.
- Kripke, S. (1980) *Naming and Necessity*. Oxford : Blackwell.
- Lawlor, K. (2001) *New Thoughts About Old Things*. New York : Garland.
- Millikan, R. (1997) Images of Identity. *Mind* 106 : 499-519.
- Millikan, R. (2000) *On Clear and Confused Ideas*. Cambridge : Cambridge University Press.
- Murez, M. (2009) Mental Files and Coreference. Part 2 of his *Self-Location Without Mental Files*, pp. 47-78. Paris : Institut Jean-Nicod.
- Peacocke, C. (forthcoming) Subjects, Consciousness and Perceptual Content. To appear in A. Coliva (ed.) *Self and Self-Knowledge*, Oxford : Oxford University Press.
- Perry, J. (2000) *The Problem of the Essential Indexical and Other Essays*. 2nd ed., Stanford: CSLI Publications.

- Perry, J. (2001a) *Knowledge, Possibility and Consciousness*. Cambridge, Mass. : MIT Press.
- Perry, J. (2001b) *Reference and Reflexivity*. Stanford : CSLI Publications.
- Perry, J. (2002) *Identity, Personal Identity, and the Self*. Indianapolis : Hackett.
- Pinillos, A. (2009) *De jure Coreference and Transitivity*. Ms.
- Pinillos, A. (forthcoming) Coreference and Meaning. To appear in *Philosophical Studies*.
- Pylyshyn, Z. (2007) *Things and Places : How the Mind Connects to the World*. Cambridge, Mass. : MIT Press/Bradford Books.
- Recanati, F. (1993) *Direct Reference*. Oxford : Blackwell.
- Recanati, F. (2010) Singular Thought : In Defence of Acquaintance. In R. Jeshion (ed.) *New Essays on Singular Thought*, pp. 141-89. Oxford : Clarendon Press.
- Reichenbach, H. (1947) *Elements of Symbolic Logic*. London : Macmillan.
- Russell B. (1903) *The Principles of Mathematics*. London: George Allen & Unwin.
- Russell B. (1918) Knowledge by Acquaintance and Knowledge by Description. In his *Mysticism and Logic and Other Essays*, pp. 209-32. London: Longmans, Green and Co.
- Russell B. (1956) *Logic and Knowledge: Essays 1901-1950*, edited by R.C. Marsh. London: George Allen & Unwin.
- Sainsbury, M. (2002) *Departing from Frege*. London : Routledge.
- Sainsbury, M. (2005) *Reference Without Referents*. Oxford : Clarendon Press.
- Schroeter, L. (2008) Why Be an Anti-Individualist ? *Philosophy and Phenomenological Research* 77 : 105-141.
- Searle, J. (1983) *Intentionality*. Cambridge : Cambridge University Press.
- Strawson, P. (1974) *Subject and Predicate in Logic and Grammar*. London : Methuen.