

On Mental Files and Unicepts:
Towards a Naturalistic Account of Direct Reference (Draft 1)

In his classic paper *On Sense and Reference*, Gottlob Frege (1947) reasons that if the content of any given singular term is its referent (as the early Bertrand Russell seems to hold¹), then it must be the case that all identity statements that relate two singular terms are uninformative, expressing merely that some individual is identical to itself. Resisting this absurd implication, Frege concludes that Russell's theory of singular terms is mistaken; the content of singular terms is given not by their referents but by the *senses* (i.e. *modes of presentation*) that they express. But what are 'senses' or 'modes of presentation' and what features must they possess to explain the informativeness of identity statements containing singular terms that express them?

Francois Recanati (2012) attributes to Frege the descriptivist claim that senses are descriptions. On this view, the content of a singular term like 'Aristotle' is not the individual that the name refers to—in this case, Aristotle—but a set of descriptions which functions to uniquely describe the individual referred to—for instance, *student of Plato* and *Tutor of Alexander the Great*. Unfortunately for Frege, this popular descriptivist interpretation² of his view has many problematic implications which are well documented by the singularist arguments of theorists such as Saul Kripke (1980), Hilary Putnam (1975), and David Kaplan (1989). Contrary to descriptivism, many identity statements containing singular terms are neither necessary nor knowable a priori; hence, 'Aristotle' still refers to Aristotle even if it turns out that he was not the student of Plato and Tutor of Alexander the Great.

In sum, Russell's Singularism implies the falsehood that all identity statements are uninformative, whereas Frege's descriptivism accounts for informative identities, but only by positing the existence of descriptive 'senses' that conflict with intuitions about the necessity and a priori knowability of statements containing singular terms. Recanati's (2012) solution to this dilemma is to claim that the 'mental equivalent of singular terms' are 'mental files'. Tracing back to theorists such as Strawson (1974) and Perry (1980), mental files are a metaphorical solution to Frege's identity problem; the content of a singular term is its referent, but the referent is determined by the term's corresponding 'file', whose

¹ Evidence of Russell's singularism is found throughout *The Principles of Mathematics* (1903). Later, in *On Denoting* (1905), Russell reacts to Frege's concerns about direct reference by adopting the more moderate view that only singular terms referring to sense data are directly referential.

² For a discussion and defense of this descriptivist reading of Frege, see Geirsson (2013).

‘merging’ or ‘linking’ to other files with different information about the same individual is taken to explain the informative value of identity statements. Mental files can be understood as an attempt to save Frege’s theory by introducing non-descriptive senses or modes of presentation.

But there are alternative ways of understanding the notion of sense which do not rely on mental files and which account for the informative value of identity statements whilst accommodating singularist intuitions. The present paper defends the claim that Millikan’s (2017, 2020) theory of ‘unicepts’ offers a more promising solution to Frege’s identity problems than does Recanati’s (2012) theory of mental files. Unicepts are “[t]he link[s] connecting stored information about the same thing together” (Millikan, 2017, pg. 7). Unicepts are formed by ‘unitrackers’, mechanisms for “for same-tracking something, for recognizing when incoming information concerns something, then linking and storing this information together as information about one and the same thing” (ibid). Unlike mental files, unicepts do not bear containment relations to the items of information they are associated with, avoiding psychologically implausible implications about working memory and the retrieval of information about the referents of singular terms. Relatedly, unicepts are neither committed to a discursive nor iconic format of representation, grounding singular thought in perception and yet avoiding Ned Block’s (2022) objection that the iconic format of perceptual representations is fundamentally distinct from the discursive format of working-memory representations. Given that unicepts are grounded in a naturalistic theory of biological function and play many of the same explanatory roles as do mental files while avoiding several objections that plague them, there is good reason to replace mental files with unicepts.

In part one, I bring the explanandum into focus by arguing that Frege’s puzzle about informative identity statements is an instance of the broader problem of explaining the mind’s capacity to recognize that multiple pieces or sources of information are about the same object. Next, I trace the history of the ‘mental files’ analogy back to Strawson’s primordial solutions to Frege’s identity puzzles.

In part two, I describe Recanati’s recent theory of mental files, explaining how it accounts for informative identity statements in terms of non-descriptive senses and how it ‘grounds’ singular thought in perceptual files. I highlight that Recanati’s theory of mental files posits discursive (i.e. non-iconic, language-like) representations, connecting the theory back to Strawson’s ‘card catalog’ analogy, one of his two original ‘pictures’ of direct reference.

In part three, I introduce Millikan’s ‘unicepts’, explaining how they account for broad question of how the mind is able to recognize multiple pieces of information as being about

the same object, thereby addressing Frege's narrower puzzle about informative identity statements. I emphasize the contrast between Recanati's Mental Files and Millikan's unicepts by pointing out that unicepts bear more resemblance to 'points' in Strawson's 'information web' analogy than to 'cards' in his 'card catalog' analogy.

Finally, in part four, I argue that because mental files and unicepts differ in several conflicting ways and yet are intended to play many of the same explanatory roles, pluralism about mental files and unicepts is incoherent. I then defend my main claim; there are at least two strong reasons to dispense with mental files and endorse unicepts. Firstly, unicepts avoid problematic psychological commitments that plague mental files—namely, commitments to a discursive format of representation and to containment relations between files and information. And secondly, unlike mental files, which rely on a normative theory of content, unicepts are grounded in a naturalistic theory about the content of mental representations.

Part One: Identity and Direct Reference

Here, I start by explaining what is at stake in the debate between singularism and descriptivism (1.1) before examining Strawson's early solution to the problem (1.2).

1.1 Clarifying the Explanandum

Gottlob Frege's (1948) classic paper opens by asking what kind of relation identity is, observing that,

[...] $a=a$ and $a=b$ are obviously statements of differing cognitive value; $a=a$ holds a priori and, according to Kant, is to be labeled analytic, while statements of the form $a=b$ often contain very valuable extensions of our knowledge and cannot always be established a priori (pg. 209).

In other words, where a and b are both singular terms—'names or signs' to use Frege's terminology—statements of the form $a=a$ are trivial and uninformative, whereas statements of the form $a=b$ are often informative. Given the informativeness of statements of the form $a=b$, there is good reason to suppose that the identity statements involving singular terms do not relate to the referents of those terms. If identity statements were to relate the referents of singular terms, then those identity statements would be uninformative and analytic, merely expressing that some individual is identical to itself.

To account for the informativeness of identity statements, Frege dispenses with the classic Russellian view that singular terms 'directly refer' to their objects, arguing instead that singular terms express 'senses' or 'modes of presentation' that refer to their objects. He writes,

It is natural, now, to think of there being connected with a sign (name, combination of words, letter), besides that to which the sign refers, which may be called the referent of the sign, also what I would like to call the sense of the sign, wherein the mode of presentation is contained (pg. 210).

Hence, the relation of identity expressed by identity statements holds not between the referents of the terms on either side of the identity, but between the ‘senses’ that those terms express. On this view, the relation between singular terms and their referents can be understood as ‘mediated’ rather than ‘direct.’ Singular terms refer, but only because the senses that they express refer. And identity statements are informative because they say of two separate senses that they refer to the same object.

But what are ‘senses’? Let us suppose—following Geirsson (2013)—that Frege thought of all senses as descriptions. Then according to Frege, the reference of any given singular term is mediated by a description. Singular terms are ‘singular’ because they each pick out a single individual who uniquely satisfies their associated description. The term ‘Aristotle’ refers to Aristotle because that is the only individual satisfying the description associated with the term—*student of Plato and Tutor of Alexander the Great*. This view—that all senses are descriptions—is widely known as descriptivism, and despite its elegant explanation of the informativeness of identity statements, descriptivism faces a number of devastating objections. In his *Naming and Necessity* lectures, Saul Kripke (1980) advances (among others) two major arguments against descriptivism—the modal argument and the epistemic argument. His arguments focus on proper names (a type of singular term), but the point generalizes to all singular terms.

The modal argument against descriptivism proceeds roughly as follows:

1. If descriptivism is true, then its not possible for the referents of singular terms to fail to satisfy the descriptions associated with those singular terms.
2. The referents of singular terms often fail to satisfy the descriptions associated with those singular terms.
3. Therefore, descriptivism is false.

Premise 1 represents an obvious consequence of descriptivism, which states that the referent of any given singular term is determined by its associated description. If some individual fails to satisfy the description associated with a singular term, then that individual is not the referent of that singular term. Hence, the claim that ‘Aristotle was not the student of Plato and Tutor of Alexander the Great’ implies a contradiction—that the student of Plato and Tutor of Alexander the Great was not the student of Plato and Tutor of Alexander the Great. But this seems false; it is possible that Aristotle was not the student of Plato and Tutor of Alexander the Great. Generalizing this point yields premise 2.

The epistemic argument against descriptivism is similar to the modal argument, the key difference being that its focus is on the way that statements involving singular terms are known, rather than the situations in which they are true or false:

1. If descriptivism is true, then all identity statements relating a singular term and its associated description are analytic or a priori truths.
2. But many identity statements relating a singular term and its associated description are not analytic or a priori truths.
3. Therefore, descriptivism is false.

Premise 1 is another consequence of descriptivism. If the content of Aristotle just is *the student of Plato and Tutor of Alexander the Great*, then ‘Aristotle is the student of Plato and Tutor of Alexander the Great’ is true simply in virtue of the meaning of ‘Aristotle’; it is known a priori. But again, this seems false. We have to study Aristotle in order to learn (a posteriori) that he was the student of Plato and Tutor of Alexander the Great. This same worry generalizes to most singular terms.

We now seem to have arrived at the following dilemma. Either singularism is true and the content of any given singular term is its referent—implying that all identity statements are uninformative—or descriptivism is true and the content of any given singular term is its associated description—implying that identity statements relating singular terms are necessary and analytic. Accepting either of these disjuncts involves rejecting deeply held intuitions about identity statements and singular terms. Either there is something wrong with how the dilemma is set up (as Strawson (1974) argues below) or we must bite the bullet by rejecting our intuitions. To make matters even worse I argue that—in addition to concerns about the contents of singular terms—the conflict between singularism and descriptivism concerns the much broader question about how minds are able to track disparate items or sources of information as being about a single object.

Frege’s puzzle about informative identity statements asks how it is possible for identity statements relating singular terms to have informative value. Another way of framing the problem asks why we have anti-substitution intuitions about singular terms occurring in sentences that are embedded in attitude contexts (e.g. contexts of believing that ..., wondering whether ..., wanting that ..., etc.). For instance, the sentences ‘Mark Twain was born in Missouri’ and ‘Samuel Clemens was born in Missouri’ express the same truth-conditions (i.e. same meaning). Thus, ‘Mark Twain’ and ‘Samuel Clemens’ can be substituted for one another without changing the content of the unembedded sentences they occur in. However, when those sentences are embedded in attitude contexts, strong anti-substitution intuitions arise. Hence, ‘Tom believes that Mark Twain was born in

Missouri' seems to express a different content than does 'Tom believes that Samuel Clemens was born in Missouri.'

As Geirsson (2013) points out, these anti-substitution intuitions can be generated with non-linguistic examples which do not involve singular terms, revealing that the problem of informative identity statements is part of a larger problem about the mind. He writes,

If we represent the same object in two or more ways, not knowing that they are representations of the same object, then we seem to invite substitution puzzles. While I will for the most part focus on names, whatever solution is provided for the problems raised by substitution needs to be informed by the nature of the problem (pg. 15).

Geirsson illustrates this point by describing Mike, the blind protagonist of the movie *The Scent of a Woman*, who relies on his sense of smell to identify people. Because she wants to rummage through Mike's room undetected, and because she knows that Mike identifies her with a particular scent, Mary masks her scent with a new perfume. Mike recognizes that someone has been in his room but does not know who. Assuming that he uses scents to represent people similarly to how most people use names to represent people, it's clear that Mike wrongly believes that the person represented by the lingering scent in his room is not Mary. Mike learns something when he realizes that the scent in his room represents Mary.

Millikan (2017) makes a related point when she discusses Helen Keller. Despite being blind and deaf, Helen Keller's thoughts are presumably about the same individuals and kinds as are the thoughts of a person with functioning eyesight and hearing. Hence, Millikan claims that "People do not need to know the same way or ways to identify either an individual or a real kind in order to think about the same and use the same word for the same. In principle, they might have entirely disjointed ways of recognizing when they are receiving information about the same individual or kind" (pg. 33). Again, we might have strong intuitions against substituting Helen Keller's beliefs and other cognitive attitudes with our own, even when they are about the same objects. Two separate beliefs can have the same content. What is at stake here is the nature of thought and its connection to the world. How does the mind recognize multiple pieces or sources of information as being about the same object? How is it that Helen Keller's thoughts can be about same objects as your thoughts, given that she recognizes and thinks about those objects in a much different way? And, relatedly, how is it that Mike's beliefs about Mary can differ, though they are both caused by scents that were caused by Mary?

Above, I raised a number of problems pertaining to singular terms and direct reference. I pointed out that either singularism is true or descriptivism is true, but both seem to have

highly problematic implications. Kripke points out that descriptivism conflicts with strong modal and epistemic intuitions about singular terms, and Frege points out that singularism conflicts with intuitions about the informativeness of identity statements. These problems are compounded by the observation that what is at stake is not merely the content and cognitive value of singular terms, but rather, the mind's ability to recognize multiple pieces or sources of information as being about the same object. In the next section (1.2), I discuss Strawson's early (1974) response to the dilemma between singularism and descriptivism about singular terms. Part two and three of this paper discuss how Strawson's metaphors are built into more substantive theories by Recanati (2012) and Millikan (2017).

1.2 Strawson's Metaphorical Explanans



Fig. 1 (above) Card-Catalog (Stock Image)



Fig. 2 (above) Web of Information on a Chalkboard (Stock Image)

Peter Strawson (1974) sketches two pictures that light the path toward a promising solution to the dilemma between singularism and descriptivism—the 'card catalog' picture and the 'web of information' picture. Ultimately, Strawson defends singularism—the content of a singular term is given by its referent, not its associated description—but denies that singularism implies that identity statements relating singular terms are uninformative. Contrary to what Frege might have assumed, Strawson's pictures suggest that the informative value of identity statements can be accounted for without invoking content.

Strawson's discussion of identity statements and direct reference focuses on one type of singular term—proper names—but the picture he presents can presumably be generalized to other types of singular terms, providing the beginning of an explanation of how the mind stores multiple pieces of information about the same object. He asks us to imagine a man as a machine that receives and stores information about the objects of which he already has some identifying knowledge. The machine contains a set of cards, each of which corresponds to a cluster of knowledge used to identify an object. Strawson explains that,

On receipt of an ordinary predication invoking one such cluster, the appropriate card is withdrawn, the new information is entered on it and the card is returned to stock.

On receipt of an ordinary relational predication invoking two such clusters, the two appropriate cards are withdrawn, cross-referring entries are made on both and both cards are returned to stock. On receipt of an identity-statement invoking two such 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 (pg. 46).

This picture suggests that the informativeness of identity statements can be explained in a way that is compatible with a directly referential account of the semantics of singular terms (i.e. singularism). Each ‘card’ in the ‘card-catalog’ of a given mind stores information about an object that the subject has some prior identifying knowledge about. Importantly, the information stored on the card need not determine the ‘referent’ of the card (i.e. the thing the card is about), as this is the function of the ‘prior identifying knowledge’. Hence, Strawson’s cards are directly referential, meaning that their referent is not determined by an associated description, but by the referents ‘identified’ by the cards. Since at least some ‘cards’ presumably correspond to singular terms, this picture explains how identity statements like ‘Mark Twain is Samuel Clemens’ are informative; two cards—the ‘Mark Twain card’ and the ‘Samuel Clemens card’—are withdrawn and discarded, and the information on each card is consolidated on a new card bearing the names of both discarded cards. This consolidation process leads to the subject gaining new beliefs about Mark Twain and Samuel Clemens.

On Strawson’s second picture—what I call the ‘web of information’ picture and Millikan (2020, pg. 259) calls the ‘wheel’ analogy—the mind is represented as a map of connected nodes, each of which corresponds to identifying knowledge about some individual. Hence, to produce a model of someone’s mind,

[...] we represent the unity of every cluster of identifying knowledge (i.e. identifying knowledge which the man regards as identifying knowledge of one and the same particular item) by a filled-in circle or dot, such as is used to represent stations on railway maps. Any name he knows which for him in suitable circumstances invokes that cluster of identifying knowledge is written adjacent to the dot. From each dot radiate lines bearing one-or-more-place predicate expressions; and these lines, with their inscriptions, represent the various propositions which the man is able to affirm, from his own knowledge, regarding the items which the appropriate cluster of identifying knowledge is knowledge of (Strawson, 1974, pg. 45).

Like the card-catalog picture, this picture also provides a way of understanding how singularism about content is compatible with informative identities. When the subject

learns that two names refer to the same individual, they presumably merge (or at least ‘link’) the clusters corresponding to those names into a single cluster.

Each of Strawson’s pictures—the card catalog picture and the web of information picture—address both (1) and (2):

- (1) *The narrow question* of identity: how a directly referential account of singular terms is compatible with the informativeness of identity statements relating singular terms.
- (2) *The wide question* of identity: how the mind stores disparate sources or pieces of information about multiple objects as being about the same object.

Both pictures address problem (1) by suggesting that the cognitive significance of identity statements can be explained without reference to content, and hence that it is possible to endorse singularism about the content of singular terms whilst denying that all identity statements relating co-referring terms are uninformative. And both pictures address problem (2) by providing metaphorical depictions of the cognitive architecture involved in storing information about specific objects. Where the card-catalog and web of information pictures differ is the type of cognitive architecture suggested by their respective metaphors.

On the card-catalog picture, cards ‘contain’ the information that the subject predicates to the individual referred to by the card. Supposing that cards are a metaphor for mental representations about individual objects, this picture suggests that mental representations ‘contain’ information about the objects they represent. For instance, the mental representation corresponding to the term ‘Aristotle’ contains the simple predicate³ ‘philosopher’ and the relational predicate ‘student of’, relating Aristotle to Plato. By contrast, on the web of information picture, nodes are (merely) ‘connected to’ the information that the subject predicates to their respective referents. Assuming nodes are a metaphor for mental representations about individual objects, the suggestion is that mental representations are ‘connected’ to information about the objects they represent. In this case, the mental representation corresponding to the term ‘Aristotle’ is *connected to* the simple predicate ‘philosopher’ and the relational predicate ‘student of’, relating Aristotle to Plato.

³ Assuming a general representational theory of mind, predicates are mental representations. Hence, as I discuss in (4.1), the card-catalog picture suggests that predicates are types of mental tokens. Because minds store information about individuals ‘inside’ of mental files, each piece of information stored corresponds to a unique tokening of a predicate type.

Recanati's (2012) mental files can be understood as a development of Strawson's card-catalog analogy into an explicit theory about the mental correlates of singular terms. By contrast, Millikan's (2017) unicepts can be understood as a development of Strawson's web of information analogy. Below, I introduce and compare both of these theories.

Part Two: Mental Files

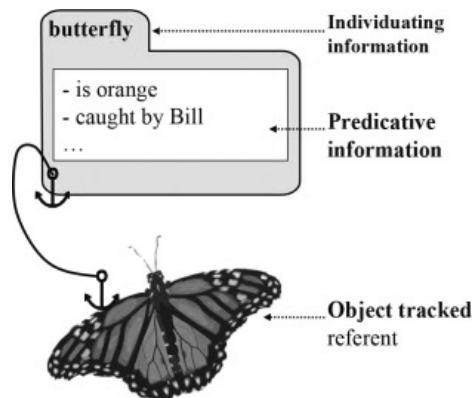


Fig. 3 (Left), Pictorial Representation of a Mental File of a Butterfly (Perner et al., 2015)

2.1 What are Mental Files?

Recanati (2012) defines 'mental files' in terms of their function, which is to store information gained in virtue of acquaintance relations that they bear to specific objects. He writes,

The role of a mental file based on a certain acquaintance relation is to store information acquired in virtue of that relation. The information in question need not be veridical; we can think of it in terms, simply, of a list of predicates which the subject takes the referent to satisfy. The referent need not actually satisfy the predicates in question, since the subject may be mistaken (pg. 37-38).

In other words, the reference of mental files is determined *relationally*, rather than *satisfactionally*⁴. This feature of mental files is what makes them directly referential; reference is not determined (or 'mediated') by descriptions. Since mental files are hypothesized to be the mental correlate of singular terms, the referents of singular terms are determined not by their respective descriptions, but by the objects that their corresponding mental files stand in some specialized relation with. But what are these specialized relations?

The relations that determine the reference of mental files are epistemically rewarding (ER) relations. Short of providing necessary and sufficient conditions, Recanati writes that ER relations "enable the subject to gain information from the objects to which he stands in

⁴ Kent Bach (1987) introduced this distinction.

these relations". He then provides examples of perceptual ER relations and mediated ER relations. Perceptual ER relations "make the perceptual flow of information possible," and 'perceptual files' "store the information gained on the basis of these short-term relations." For instance, the subject who holds a ball and looks at it bears a perceptual relation to the ball, storing the information gained in virtue of that relation in a mental file. By contrast, mediated ER relations are based on Kripkean 'communicative chains,' and the files based on them "[...] enable the subject to (through communication) gain information from the object at the other end of the communicative chain" (pg. 35). For instance, 'Aristotle' refers (if at all) to Aristotle because the corresponding mental file bears a relation to him through a long communicative chain. In sum, mental files store information gained in virtue of ER relations of which there are two paradigmatic types—perceptual ER relations and mediated ER relations. The object that a given file stands in an ER relation with is its referent. Additionally, ER relations serve to individuate files; for each mental file there corresponds a unique ER relation. Call this 'the uniqueness constraint'.

The uniqueness constraint gives rise to a problem; information about individual objects must be retained despite changes in the relations that subjects stand in with their environment. For instance, consider directly referential indexicals⁵ like 'yesterday' and 'today' and their corresponding mental files. If the 'today file' refers to today because it stands in an ER relation with today, then when tomorrow comes, the 'today file' will fail to refer to today. Anticipating this problem, Recanati claims that files are 'dynamic' and can be 'converted' into one another. Temporal indexicals are systematically converted as time passes so that tomorrow the information stored in the today file will be stored in the yesterday file (Recanati, 2012, pg. 81).

Consider how mental files build Strawson's card-catalog analogy into an explicit theory of the mental correlates of singular terms. Like cards in the card-catalog analogy, and unlike nodes in the web-of-information analogy, files are taken to contain or store information. Specifically, files are taken to contain or store information acquired in virtue of the ER relations they stand in. This way of thinking about mental files moves beyond mere analogy towards explicit theory; mental files are whatever mental capacity or representation that functions to contain or store information acquired through ER relations. But to explicate this further it is necessary to characterize the containment relation that stands between files and information. What does it mean to say that a file 'contains' information?

⁵ Kaplan (1989) argues that indexicals are directly referential, referring to the same individual in every world they are evaluated in.

There are two key features that seem to characterize the containment relation between files and information. The first feature relates to information insulation or clustering. Describing this aspect of containment, Recanati writes,

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 licenses the integration and inferential exploitation of the information in question (pg. 42).

So, to say that a file contains information is to say that the information acquired in virtue of that file's ER relation can be integrated and inferentially exploited. For instance, before learning that Mark Twain is Samuel Clemens, the subject is not licensed to infer that the properties applying to Mark Twain also apply to Samuel Clemens. But when the subject learns that Mark Twain is Samuel Clemens, the information in each corresponding file can be exploited for inference, allowing the subject to infer that (for instance) Samuel Clemens wrote *The Adventures of Tom Sawyer* and worked on a steamboat.

The second feature characterizing the containment relation is its asymmetry. Where A is a file and B is a predicate, if A contains B, then B does not contain A. It makes sense to say that files contain predicates, but not to say that predicates contain files. As Recanati writes, "[...] files contain predicates [and] [a]ll the predicates in a given file record information concerning the individual the file is about (the referent of the file)" (pg. 50). A notable consequence of this view—where files contain predicates and containment is asymmetric—is that the predicates stored in each file are unique tokens of general predicate types. If the subject believes that *Jill is beautiful* and that *John is beautiful*, then the subject stores two tokens of the predicate *is beautiful*—one token in their Jill file and the other in their John file.

2.2 How do Mental Files Solve Identity Problems?

Mental files address how a directly referential account of singular terms is compatible with the informativeness of identity statements containing them. As discussed above, mental files account for the informativeness of identity statements in terms of the cognitive significance (rather than content) of singular terms. Becoming informed of an identity involves linking the mental files corresponding to each relata of that identity. This allows the information in each file to be integrated and exploited for inference.

Mental files also account for how the mind stores disparate sources or pieces of information about multiple objects as being about the same object. Answering this general question requires tending to how the mind can recognize and track objects of perception

and later ‘refer’ to them in thought. Linking singular thought and perception, Recanati (2020)⁶ writes that,

[...] the object tracking system which exists in perception is used throughout cognition — even in high-level cognition, e.g. in thought about abstract objects. The units of the system — the files — acquire new functions when they are recruited in this way to do extra jobs in thought (pg. 212).

Recanati claims that the mental files of perception are evolutionarily ‘basic’ or ‘primary’, as compared to the ‘derived’ mental files of thought. Hence,

We start with referential devices which are already in place in perception (as well as in infant cognition). These devices are encoded in language, but language makes it possible to do things with them that we can’t do in perception (or that infants can’t do) (pg. 211-212).

So mental files are not only the constituents of thought, but also the means by which the mind stores information about objects of perception.

Part Three: Unicepts

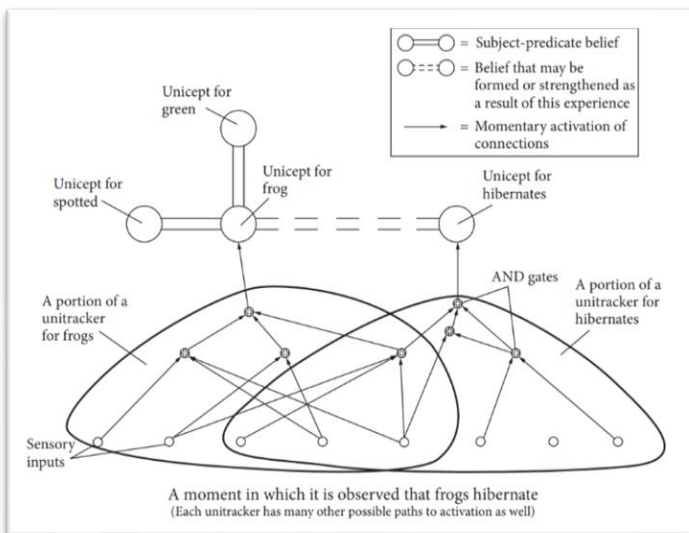


Fig. 4 (left),
Diagrammatic
representation
of unicepts
(Millikan, 2017,
pg. 50)

Below, I explain Millikan’s unicepts by explicitly contrasting them with Recanati’s mental files.

3.1 What are Unicepts?

Similar to Recanati’s mental files, Millikan’s (2017) ‘unicepts’ are defined functionally, as “[t]he link connecting stored information about the same thing.” Unicepts work in tandem

⁶ See also, Murez and Recanati (2016).

with ‘unitrackers,’ which are “mechanism[s] or facult[ies] for same-tracking something, for recognizing when incoming information concerns it, then linking and storing this information together as information about one and the same thing (pg. 7).” Together, unicepts and unitrackers allow the subject to,

[...] recognize Obama again from the front or the back or in a newspaper photograph or by his voice, to recognize tiger again given different views or sounds or kinds of spoor, to recognize anger again given different facial or bodily expressions or sounds or behaviors, to recognize zero degrees again given different ways of measuring temperature, to recognize multiple sclerosis again given different symptoms, to recognize information coming in about Venus, whether called “the morning star,” “the evening star,” “Hesperus,” or “Phosphorus (pg. 7).”

In sum, unicepts track when two singular terms pick out the same object, or when two different perceptual states represent the same object. So unicepts play a similar role as do mental files; both are theoretical posits that enter into explanations of how the mind identifies and stores information about particular objects.

Millikan’s theory of unicepts can be understood as a development of Strawson’s ‘web-of-information’ which pictures the mind not as a set of cards or files that contain information, but as a set of nodes representing pieces of identifying knowledge and lines that connect related nodes together. Explicating this account of cognitive architecture, Millikan further describes the function of unicepts and their partnered unitrackers. Unicepts act as the glue that unitrackers use to connect separate items of knowledge that are about the same object. A unicept “holds [items of knowledge] together so they can work together as, for example, a middle term in inference.” Hence,

“Uni” is for *one* and “cept” is from *capere, to take or to hold*. The unicept takes various items of information about one and the same thing and holds them together so that they can be used together. The referent of a unicept is its unitracker’s target, what the unitracker has been designed to track (Millikan, 2017, pg. 44).

The connections formed by unicepts license inferences between nodes representing items of knowledge. For instance, the unicept which you use to determine whether various types of incoming information are about ‘dogs’ connects knowledge about mammals, pets, furriness, etc. Hence, if you gather information about something that is a pet, a mammal, and furry, then you will be more likely to infer that the thing you are gathering information about is a dog.

Though Millikan is agnostic about how unicepts and unitrackers are physical instantiated in brains, she suggests that ‘unitrackers’ are housed in “[...] a neural network devoted to recognizing affording and/or factual information about some one particular thing,

collecting it together for combined use” (pg. 44). She also offers an explicitly neural definition of unicepts in the glossary, writing that a unicept is a “[...] neural node that helps in storing factual or procedural knowledge through its connections with other unicepts or with behavior controllers” (pg. 225). Picturing how unicepts and unitrackers are instantiated in actual networks helps to understand how they work (see figure 4).

For present purposes, there are at least two important differences between mental files and unicepts. Firstly, unlike mental files, unicepts do not ‘contain’ information about their referents. Rather, unicepts are ‘connected’ with information about their referents. This means that there is no substantial difference between how individuals and kinds are represented. In other words,

[...] no particular way of being able to identify or characterize a thing is required for understanding a name of an individual or real kind, or for being able to think of the individual or kind, so long as one has some way of identifying it (pg. 34).

Recall that files contain predicates; each file contains unique instances of all of the predicates attributed to the referent of that file. This is not the case with unicepts, which do not ‘contain’ but rather ‘connect’ information together, weighing some connections more or less than others. Your unicept representing Aristotle is likely inferentially connected to the unicept representing philosophers, but neither has a special ontological status, and neither ‘contains’ the other. Both terms—Aristotle and philosopher—are directly referential. Indeed, Millikan claims that all extensional terms are directly referential, picking out real kinds (i.e. ‘property clusters’) in the world.

The second important difference between files and unicepts is that, unlike files, the referent/content of a unicept is determined not by a particular type of ER relation to the object, but rather, by the object that the unicept has been designed to track. This feature of unicepts is what grounds Millikan’s theory in biology and, more specifically, in the selectional history of the consumer of the unicept. According to Millikan, the content of mental representations—and hence, unicepts—is determined by how the consumer of those representations exploits them. Clarifying her view, Millikan (2009) writes,

the producer produces a sign that will be true or satisfied only if it maps on to some affair or affairs [...] in the world in accordance with certain ‘semantic’ rules. These are rules of correspondence between signs and world affairs that have been instantiated in the past when the consumer and producer or their ancestors have succeeded in performing their cooperative functions (pg. 397-398).

In other words, ‘correspondence rules’ (or ‘normal mapping rules’ to use Millikan’s (1989) terminology) are rules which describe some relation between representations and represented affairs, and which are necessary for providing a ‘normal explanation’ of how

the producer and consumer have historically cooperated to perform their proper function. The content of a representation is ultimately determined by the proper function of the consumer; the content of a particular representation is given by a correspondence rule between the representation and some state of affairs which has held in instances where the consumer has historically succeeded in normal performance of its proper function.

This theory about how representations—and hence, unicepts—get their content, is radically different from Recanati’s story about how mental files get their content. The content of a mental file is just the individual it bears an ER relation with, and as stated by the uniqueness condition, each mental file corresponds to a single unique ER relation. This is not the case with Millikan’s unicepts, whose content is determined by the correspondence between it and external world affairs that has been instantiated when the unicept contributed to the performance of the consumer’s proper function.

Part Four: Towards a Naturalistic Account of Direct Reference

Recall the *narrow* and *broad* questions of identity from (1.1):

The narrow question of identity: how a directly referential account of singular terms is compatible with the informativeness of identity statements relating singular terms.

The broad question of identity: how the mind stores disparate sources or pieces of information about multiple objects as being about the same object.

Below, I show that though mental files address the *narrow question* of identity, their account of the *broad question* of identity faces several serious objections. Unicepts answer the narrow and broad questions of identity whilst avoiding the serious objections to mental files.

4.1 Objections to Mental Files

4.1.1 Objection 1: Storage Bins and Working Memory

The relation that unicepts bear to stored information about their referents is significantly different from the relation between mental files and the stored information about their referents. In (2.1), I emphasized that the containment relation that files bear to predicates is asymmetric. In (3.1), I pointed out that the relation that unicepts bear to information that the subject stores about the referents of those unicepts is symmetric—this follows from the fact that individuals and kinds are represented in the same way, and neither contains the other. Recognizing this difference, Millikan (2020) emphasizes that “the files model requires the sameness of subjects and of their properties to be marked in different ways; numerical identity for subject terms, duplicates marking for predicate terms,” and that,

The job of a unicept is not like that of a folder. It cannot be just a storage vessel storing knowledge of the predicates of one subject. [...] In this way a vast amount would be represented directly that might otherwise require hunting through huge storage bins looking for relevant premises for inference (Millikan, 2020, pg. 259).

In other words, the asymmetric containment relation that files bear to stored information has implausible implications for working memory. In Millikan's words, the theory of mental files implies that retrieving premises for inference would require searching through 'huge storage bins'.

To bring out this storage bin objection more clearly, consider what mental files seem to imply about how subjects retrieve information from their memory. For a subject to retrieve information about a given kind or category of thing, they would have to find each stored instance of that kind. This follows from the fact that predicates are stored as tokens inside files. So, for instance, if the subject tries to recall all of the philosophers they know, they would have to locate each token of their philosopher predicate, which would in turn require searching through every file and checking whether each file stores the philosopher predicate. This procedure would be incredibly inefficient and would make it very difficult for the subject to recall sets of experiences that have similar characteristics, or that fall under a common type.

Unicepts do not face this problem because the relation that they bear to information about their targets is symmetric, and individuals are represented in the same way as kinds. The David Hume file does not 'store' or 'contain' the philosopher predicate. Rather, there is an inferential connection (undergirded by a neural connection) between the unicept representing Hume and the unicept representing philosophers. The subject recalls the philosophers they know by activating their philosopher unicept, which is inferentially connected to each unicept corresponding to individual philosophers.

4.1.2 Objection 2: Iconic and Discursive Representational Formats

In psychology and the philosophy of perception there is lively debate about the format of perceptual representations. A main point of contention is whether perceptual representations have the same representational format as do the representations of thought and working memory (Quilty-Dunn, 2016, 2019, & 2020). In a recent paper entitled *Let's Get Rid of the Concept of an Object File*, Block (2023) draws from numerous case studies to convincingly argue that the perceptual representations have a fundamentally different type of format than do working memory representations; perceptual representations are 'iconic' whereas working memory representations are 'discursive'. This raises a significant problem for Recanati's theory of Mental Files, which attempts to ground singular thought in perception. Mental files are inherently discursive. Hence, if Block's

thesis is true, then mental files cannot account for perceptual representations. By contrast, unicepts are not committed to a particular format of representation, and hence, are not constrained by this limitation.

What is the difference between iconic and discursive representations and why are mental files necessarily discursive? Quilty-Dunn (2019) uses an image of Bob Dylan wearing a checkered shirt as a paradigmatic example of an iconic representation. By contrast, the sentence “Bob Dylan is wearing a checkered shirt” is a paradigmatic example of a discursive (i.e. language-like) representation. Though Quilty-Dunn (2019) hesitates to provide necessary and sufficient conditions for discursive and iconic representations, he does identify some strong points of comparison, writing that

[...] parts of icons correspond to parts of what they represent, while this need not be true of discursive representations. Typically, distance relations between parts of icons correspond to distance relations in what is represented [...] and] icons represent holistically, i.e., parts of icons encode various properties at once. Icons do not represent by means of a canonical decomposition into constituents that stand separately for distinct individuals and distinct features, while discursive representations may and ordinarily do have such a canonical decomposition (Pg. 5-6).

Under this description of the distinction between iconic and discursive representations, mental files are necessarily discursive because they represent by means of decomposition into the individuals they bear ER relations with and the predicates they store. Furthermore, Mental files do not contain parts whose relations correspond to distance relations in their referents.

What is Block’s (2023) reason for thinking that perceptual representations are ‘iconic’ whereas working memory representations are ‘discursive’? Block points to a range of experimental cases where information that is clearly present in perception seems to degrade or change format once it is carried over into working memory. For example, he discusses visual phenomena like crowding and certain grouping effects, where in immediate perception you can effortlessly register fine-grained spatial relations or ensemble properties (like the average orientation of a set of lines), but when you have to hold that same information in working memory, people tend to rely on more abstracted summaries or lose access to those detailed spatial interactions altogether. He also draws on change-detection and delayed-report tasks showing that subjects can initially experience rich, high-resolution scenes, yet when asked to retain and report specific features after a delay, their responses look more like they are based on a limited set of

discrete, concept-like representations (e.g., “a red vertical bar on the left”) rather than anything like a detailed mental image. Another strand of evidence comes from studies where perceptual computations—like certain visual illusions or context-sensitive effects—simply don’t carry over into working memory: once the stimulus is gone, the stored representation no longer behaves like something embedded in a perceptual field but instead like something that has been recoded into a more symbolic format. Taken together, these cases suggest (at least to Block) that perception itself really does operate in this rich, image-like, iconic way, while working memory involves a kind of translation into a discursive code that strips away a lot of that fine-grained structure and reorganizes the information into something more like a sentence or description that can be used in reasoning and report.

4.2 Conclusion

As argued in part one, mental files and unicepts are theoretical posits aimed at explaining how singularism about the content of singular terms is compatible with the informativeness of identity statements, and how the mind is capable of storing disparate sources or pieces of information about multiple objects as being about the same object. Recanati’s (2012) mental files build from Strawson’s (1974) card-catalog analogy. The mental correlates of files ‘contain’ information gained in virtue of an epistemically rewarding (ER) relations to the respective referents of those files. By contrast, Millikan’s (2017) unicepts build from Strawson’s (1974) web of information analogy. Unicepts are ‘connected’ to information about the individuals and kinds which of which it is their proper function to track.

Since mental files and unicepts play a similar explanatory role, and since they make contradictory commitments about the mechanisms that play that role, it is incoherent to be a pluralist about mental files and unicepts. Both mental files and unicepts aim to explain how it is that the mind recognizes and stores information about individual objects. But mental files bear containment relations with information they store in the form of predicate tokens, whereas unicepts are merely inferentially connected to other unicepts, and there is no special difference between the way predicates and individuals are represented. Furthermore, the content of a unicept is determined by the proper function of the producer and consumer of that unicept, whereas the content of a mental file is determined by the object it bears an ER relation to.

Given that it is incoherent to be a pluralist about mental files and unicepts, we must find a way to adjudicate between the two. In part four, I argued that unicepts avoid at least two of the serious objections that plague mental files. Unlike mental files, unicepts do not imply

the implausible view that we retrieve information about kinds by searching through large sets of files or ‘storage bins’, as Millikan puts it. Next, I argued that unlike mental files, unicepts are not committed to a discursive representational format, implying that if perceptual representations are non-discursive, unicepts may still be able to account for them.

I have not argued for Millikan’s view of unicepts, but rather, shown that unicepts avoid many of the theoretical problems facing mental files. Millikan’s theory of unicepts faces many objections not addressed here, some of which are addressed in Shea’s (2018, section 1.5) *Representations in Cognitive Science*. Shea argues that we should dispense with Millikan’s notion of proper functions, replacing it with ‘task functions.’ Furthermore, Shea expresses warranted skepticism about Millikan’s distinction between producers and consumers. These worries do not seem fatal to Millikan’s theory of unicepts, but developing the theory would require addressing them.

Citations:

- Bach, Kent (1994). *Thought and Reference*. Oxford, GB: Clarendon Press.
- Block, Ned (2023). Let's get rid of the concept of an object file. In Brian McLaughlin & Jonathan Cohen, *Contemporary Debates in Philosophy of Mind*, Second Edition. Wiley.
- Frege, Gottlob (1948). Sense and reference. *Philosophical Review* 57.
- Geirsson, Heimir (2013). *Philosophy of language and webs of information*. New York: Routledge.
- Jin, Zhengxi (2026). Embedding mental files in the world. *Mind and Language* 42 (2):1-21.
- Kaplan, David (1989). Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals. In Joseph Almog, John Perry & Howard Wettstein, *Themes From Kaplan*. New York: Oxford University Press.
- Kripke, Saul Aaron (1980). *Naming and Necessity: Lectures Given to the Princeton University Philosophy Colloquium*. Cambridge, MA: Harvard University Press. Edited by Darragh Byrne & Max Kölbel.
- Millikan, Ruth (2009). Biosemantics. In Ansgar Beckermann, Brian P. McLaughlin & Sven Walter, *The Oxford Handbook of Philosophy of Mind*. New York: Oxford University Press.
- Millikan, Ruth Garrett (1989). Biosemantics. *Journal of Philosophy* 86 (6):281-97.
- Millikan, Ruth Garrett (2017). *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford, United Kingdom: Oxford University Press.
- Millikan, Ruth Garrett (2020). Embedding Language in the World. In Goodman, Rachel ; Genone, James & Kroll, Nick. *Singular Thought and Mental Files*. New York, NY, United States of America: Oxford University Press.
- Perner, Josef; Huemer, Michael & Leahy, Brian (2015). Mental files and belief: A cognitive theory of how children represent belief and its intentionality. *Cognition* 145.
- Perry, John (1980). A Problem About Continued Belief. *Pacific Philosophical Quarterly* 61.
- Putnam, Hilary (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science* 7.
- Quilty-Dunn, Jake (2016). Iconicity and the Format of Perception. *Journal of Consciousness Studies* 23 (3-4):255-263.
- Quilty-Dunn, Jake (2019). Is Iconic Memory Iconic? *Philosophy and Phenomenology Research* 101.
- Quilty-Dunn, Jake (2019). Perceptual Pluralism. *Noûs* 54.
- Recanati, Francois (2012). *Mental Files*. Oxford, GB: Oxford University Press.
- Recanati, François (2013). Mental Files: Replies to my Critics. *Disputatio* 5.
- Strawson, P. F. (1974). *Subject and Predicate in Logic and Grammar*. Burlington, VT: Routledge.