

Confusion is corruptive belief in false identity

Elmar Unnsteinsson

Postprint. Please cite published version:
Canadian Journal of Philosophy (2016) 46(2): 204–227
doi:10.1080/00455091.2016.1153994

Abstract

Speakers are confused about identity if they mistake one thing for two or two things for one. I present two plausible models of confusion, the Frege model and the Millikan model. I show how a prominent objection to Fregean models fails and argue that confusion consists in having false implicit beliefs involving the identity relation. Further, I argue that confused identity has characteristic corruptive effects on singular cognition and on the proper function of singular terms in linguistic communication.

1 Introduction

What is it to be confused about the identity of an object? People often confuse things. They think one thing is two or two things are one. Also, ever since Frege argued for the distinction between sense and reference, this type of mistake has been the driving force behind a host of influential arguments and theories in philosophy of language and mind. Until recently, however, theorists did not pay much attention to the metaphysics of confused mental states. The point of this paper is to argue for and present a new metaphysical theory of confusion, called the 'Frege model.'

In §2 I describe two independently plausible models of confused identity. The first is the Frege model already mentioned and the second is called the 'Millikan model.' In §3 I present an argument against the Frege model—called *the objection from unavailable*

*(✉) elmar.geir@gmail.com

*I would like to thank Hrafn Ásgeirsson, Daniel Deasy, Michael Devitt, Cressida Gaukroger, Daniel Harris, Rachel McKinney, Eliot Michaelson, Stephen Neale, Gary Ostertag, Ben Phillips, David Plunkett, Jesse Prinz, François Recanati, and David Rosenthal for comments and suggestions. I also want to thank the responsible editor and anonymous reviewers for this journal.

representation—and show in some detail why it fails. In addition, I argue that the model is superior both in terms of descriptive and explanatory adequacy.

Interestingly, however, many of the insights of the Millikan model can, and should, be incorporated into our final theory of confused identity. As Ruth Millikan has argued extensively, confused identity is a type of corruption of the *proper function* of our capacity to reidentify particulars. In §4, I explain how Millikan's notion of proper functions can be applied to the case of confused identity, first in thought and then in communication. Particularly, I show how confusion essentially disrupts the proper function of singular terms in natural language, even if the Frege model is correct.

2 Two models of the mental state of confusion

Confused identity is a mental state of two basic types. *Combinatory confusion* is when an agent takes two things to be a single thing. *Separatory confusion* is when an agent takes a single thing to be two things. In principle an agent can be confused by taking n things to be m things for any distinct natural numbers n and m. For convenience, I focus only on the more simple cases.[1] A dog who holds a bone in his mouth and takes its reflection in a pond to be a different bone suffers, albeit momentarily, from separatory confusion. Another dog who chases two similar rabbits, only ever seeing one of them at any given moment, might suffer from combinatory confusion. If the rabbits switch roles behind a rock, appearing to the dog as if a single running rabbit disappears from sight only for a moment, he seems to have some kind of mental representation that is supposed to help him track a single rabbit but really tracks two rabbits.

2.1 Frege vs. Millikan

How exactly should confusions like those attributed to the two dogs be explained and modeled? There seem to be two plausible ways to go here (cf. Lawlor 2005).

[1] Note that this is not supposed to be an analysis of the meaning of the word 'confusion' in English. The characterization is an attempt to describe what confusion itself consists in. It seems like the *word* 'confusion' is partly negative and is not properly applied, in English, unless the agent is taken to bear some responsibility for their sorry epistemic state. This would arbitrarily disqualify many cases that a general theory of confusion ought to capture. Further, theorists who have written specifically about confusion tend to think only of combinatory confusion (cf. Camp 2002; Lawlor 2007). This may simply reflect common beliefs about the etymology and morphology of the lexical item 'confusion' in English. Translations into other languages do not always give rise to the same tendency (German 'Verwirrung,' Icelandic 'ruglingur,' etc.). But of course there is a strong tradition of discussing separatory confusions under the head of Frege's puzzle.

1. *Frege model.* According to this model confusion is a mental state that consists in someone's having identity beliefs that are false. It is on this basis that Fregeans have postulated *modes of presentation* or *senses* as semantic values for linguistic or mental representations. A mode of presentation can be thought of as an identifying condition that is specified by a description. Most famously, Frege proposed that someone might think of Venus via the identifying condition of *being the morning star* or of *being the evening star*. And it is not necessary that they realize that these conditions determine the very same object. Modes of presentation, however, are an optional feature of the Frege model. It is 'Fregean' only because of its role in some of the most influential arguments for postulating such modes. But let us state the model more precisely.

   Frege model of confusion

   Speaker A suffers from confused identity at time t iff there are singular terms a and b and propositional attitudes V and W that A explicitly holds at t, such that (i) A Vs *that Fa*, (ii) A Ws *that Gb*, and only one of (iii–iv) is true; (iii) $a = b$ and A believes that $a \neq b$ (or lacks the belief that $a = b$), (iv) $a \neq b$ and A believes that $a = b$ (or lacks the belief that $a \neq b$).[2]

   'F' and 'G' are arbitrary predicate expressions of the speaker's language, standing for any two properties A can cognitively represent. They can, for example, be identical or mutually exclusive. Similarly, V and W can be, for example, the same (e.g. belief), completely different (e.g. hope and fear respectively), or mutually exclusive (e.g. belief and disbelief respectively). Finally, the definition employs the terminology of explicit propositional attitudes 'at a time.' What this means is just that there is some specific time span in which A has two explicit mental states whose contents represent a and b. In many cases, the two states can be collapsed into a single mental state whose content represents a relation holding between a and b, i.e. (i–ii) can be: A Vs *that Rab*. But the terminology is used here, firstly, to exclude cases where the mental states occur at two very distant points in time and, secondly, to contrast them with the identity beliefs (iii–iv) that A may have implicitly.

[2] To be clear, the definition is intended in such a manner that A Vs or Ws *the proposition* associated with the expression '*that Fa*.' That is, A does not stand in propositional attitude relations to particular *sentence* types or tokens containing singular terms a and b, but only to the propositions, whatever they turn out to be, associated, whatever that relation really is, with the sentences. Also, by 'singular term' I mean any expression that the speaker intends to stand for an object, it need not be a referring expression; it can be a proper name, demonstrative, indexical or a definite description used referentially. See also §4.1 below. Note, finally, that, in the terminology introduced in the next section, the Frege model is couched entirely in the t-language, not in the c-language.

2. *Millikan model.* Ruth Millikan (1994, 1997, 2000) develops a subtle theory of confused thoughts. On her view, confused identity cannot consist in having a false belief about an identity, rather, it is an error of its own kind (1994: 96, 2000: 173). In the 'central cases,' misidentifying the object of one's thought is 'an act that muddies the thought involved, corrupting the inner representational system' (1994: 75). This is taken to mean that confused representations involve concepts that are themselves confused. Here is how I propose, in more detail, to flesh out the Millikan model. To be clear, the formulation is influenced by Millikan, but it is not intended as a perfect exposition of her broader view. I want to focus on the model in its own right. A confused singular representation is one where the number of objects it actually applies to differs from the number of objects it is, in effect, taken to apply to by its user. Consider the dog chasing two rabbits. He, let us suppose, possesses a singular representation or quasi-representation C_{R} such that it actually applies to a spatially discontinuous object that consists of two individual rabbits, Rabbit *A* and Rabbit *B*. The representation C_{R} is corrupt, however, since its proper function is to apply to a single spatially continuous object and, one may add, the dog seems to quasi-believe that C_{R} really performs this function.[3]

   Now consider separatory confusion. Suppose that the dog carrying the bone in his mouth has a singular representation or quasi-representation C_{B} of that particular bone. On the Millikan model, as soon as the dog confuses the reflection of the C_{B}-bone in the pond for some *different* bone, his tokening of the C_{B}-representation applies to nothing. It applies to nothing because there is no object which consists of *that-bone-in-the-mouth* without consisting of *that-pond-reflected-bone* at the same time. So the C_{B}-representation actually applies to no object, although its proper function is to apply to a single spatially continuous object.

It should be noted here, in particular, that the Frege model implies that cases that David Kaplan (1968, 2013) and others have described in terms of suspended judgment are in fact cases of confusion. This may seem implausible but the following argument indicates otherwise.

Here is a typical scenario in which a speaker (call her 'Lois Lane #1') satisfies the Fregean definition of confusion. Lois #1 believes that Superman can fly. She also believes that Clark Kent cannot fly. As a matter of fact (assuming this is fact rather than fiction), Clark Kent is Superman. And, finally, Lois #1 does not believe that Clark Kent is Superman. Now consider a case that is less typical (relative to the massive literature

[3] The notion of 'proper function' is explained in §4 below.

on paradigm confusion cases). Another Lois—'Lois #2'—is likewise confused, on the proposed definition, if she hopes that Superman saves her and wonders whether Clark Kent has read Shakespeare while not believing that Superman is Clark Kent, possibly even suspending belief.

Many philosophers have argued that someone like Lois #1 is confused but not *irrational*. On one account of directly referential singular terms, however, it would follow that Lois #1 is being irrational, because she has inconsistent beliefs about Superman/Clark Kent. Let us call this account 'Millianism.' According to Millianism the content of a belief expressed by referring directly to object o will have an o-dependent truth condition.

What about Lois #2? On my definition she is also confused about the identity of Superman/Clark Kent. Here is my reason for taking this view. At time t, let us assume, Lois #2 tokens the hope that Superman saves her and the wondering-whether Clark Kent reads Shakespeare. But it is not the case at t that Lois #2 believes explicitly that

(1) $\lambda x\langle$she tokened the hope that x saves her and she tokened the wondering-whether x reads Shakespeare$\rangle\, o$.

Where $o =$ Superman $=$ Clark Kent. Still, this very belief can be *derived* from the set of her beliefs at t and is then, arguably, one type of implicit attitude. The derivation takes a very strong form of Millianism for granted and, plausibly, assumes that conscious, reflective individuals can reliably form true higher-order beliefs about their own propositional attitudes. Using only variables, at t, Lois #2 tokens a V-ing of Fa and a W-ing of Gb. Assuming that she knows some symbolic logic, she can then always derive the proposition that, at t, she believes that $\lambda x[\lambda y\langle$she tokened a V-ing of Fx and a W-ing of $Gy\rangle\, a]\, b$. And since in this case it is true *that* $a = b$ the imagined Millian theorist must predict that there is only a single object o to which Lois #2 refers in deriving the λ-sentence. Thus, given these assumptions, the logician Lois #2 must derivatively and implicitly believe (1) or something very much like it.

Lois #2 does not explicitly believe the negation of (1) at t, yet this can also be derived from her beliefs at t. To make the point briefly, if one were to put (1) to her in the form of a question, she would respond negatively if she responds honestly and in accordance with her actual beliefs. Here is the question I have in mind: 'Do you believe that having the property of being an x such that you have a token hope that x saves you and you have a token wonder whether x reads Shakespeare applies to a single object o, o being either the object to which you refer by 'Superman' or the object to which you refer by 'Clark Kent'?' Again, assuming that she knows some logic, Lois #2 ought to understand such contorted questions perfectly. It seems, then, that Lois #2 implicitly believes a contradiction.

Thus, if a two-tiered Fregean semantics can be motivated by taking Lois #1-type cases as data it should be equally well motivated by Lois #2-type cases. This contention is one reason for making the Frege model of confusion so general. The definition is the most general formulation of the reason many theorists believe in Fregean senses or something playing the role of such senses, such as Salmon's (1986) pragmatic 'guises.'[4] This also has the surprising consequence that suspending belief in the identity of an object about which one has explicit propositional attitudes at a specific time may amount to being confused.

3 The objection from unavailable representation

Joseph Camp (2002: 33) and Millikan (1994: 97) have voiced similar arguments against something like the Frege model, claiming that the relevant representations of identity are unavailable to the confused thinker.[5] They state the objection in terms of combinatory confusion, which I will do here as well since it is more intuitive. But the application to separatory confusion is explained at the end of the section. Following Millikan (ibid.), imagine that Bill and Biff are identical twins and *A*, their acquaintance, has met both of them on multiple occasions. But *A* has never seen both of them *at the same time* and has in fact enjoyed Bill's company for exactly the same amount of time as Biff's. Neither person commands a dominant role over the other in *A*'s mental economy.[6] To make things easier for us (the theorists), assume this is some sort of ploy, and the twins have intentionally led *A* to believe that they are a single person called 'Phil.'

[4]If it is right that Lois #2 has inconsistent beliefs then this would be avoided as soon as one adopted a Fregean theory of content. According to a typical theory of this sort the proposition believed by Lois #2 can only contain modes of presentation (call them 'MPs'), never the objects themselves. The derivation of (1) fails on Fregean assumptions because there are, we can suppose, only two modes of presentation that Lois #2 can possibly associate with the occurrence of '*o*' at the end of (1). First, there is the MP she associates with the name 'Clark Kent.' At *t*, however, Lois #2 did not token a hope the content of which had her mode of presentation of Clark Kent as a component, thus this cannot be the MP associated with '*o*' in (1). Second, there is the 'Superman'-MP. At *t*, Lois #2 did not token a wondering whether the content of which had the 'Superman'-MP as a component, thus this cannot be the MP associated with '*o*' in (1). Therefore, there is no Fregean interpretation of the λ-expression in (1) on which the claim made by it can be true and thus (1) cannot be derived. Therefore, Fregeanism saves Lois #2 from irrationality because only the negation of (1) can actually be derived from the set of her beliefs at *t*.

[5]Ruth Marcus (1981, 1983, 1990) appears to toe the same line when she argues that it is impossible to believe what is metaphysically impossible (cf. Richard 2013).

[6]Camp (2002: ch. 3) develops, at some length, the example of Fred and the ant colony, to make sure that it is possible to have cases of such pure and complete combinatory confusion. His argument is compelling. See also Lawlor (2007: 162).

If we go along with the popular 'mental file' metaphor, *A* puts all the information gathered from any epistemic encounter with Bill or Biff into the very same mental file, labelled with a single 'tag'—represented linguistically as 'Phil.'[7] Here, we need a working distinction between the language, or other system of mental representation, used by the confused person and the language used by the theorist in describing and explaining the confused mental state. Call the theorists' idiolect 't-language' and the confused person's idiolect 'c-language.' The c-language contains singular terms that correspond to mental representations which either (i) mentally combine two objects into one or, together with another representation, (ii) mentally separate one object into two. Accordingly, we can call such singular terms in the c-language *confused* (or 'c-terms' for short).

For sake of argument, then, let us assume that *A*'s c-language contains no more than one singular term or mental representation of Bill/Biff, because all the relevant information is contained in but a single mental repository with a single tag. According to the objection from unavailable representation this scenario makes it impossible, in any reasonable sense, for *A* to have two distinct representations, one of Bill and the other of Biff. But then the Frege model cannot be right, at least not as a model of someone like *A*, as it clearly requires this to be possible. If someone like *A* exists, the model must be wrong.

Camp and Millikan state the objection, then, as a *reductio* of (2).

(2) *A* believes that Bill = Biff.

In *A*'s c-idiolect there is one relevant singular term 'Phil$_c$,' and, according to Camp and Millikan, there are only three possible ways of assigning reference to it. It must refer only to Bill or only to Biff or to Bill/Biff. Double brackets '$[\![\ ]\!]$' represent functions from expressions to their referents or semantic values.

(3) $[\![\text{Phil}_c]\!] = \text{Bill}$

(4) $[\![\text{Phil}_c]\!] = \text{Biff}$

(5) $[\![\text{Phil}_c]\!] = \text{Bill/Biff}$

The last interpretation (5) is construed, according to the Millikan model, such that Bill/Biff is a spatially discontinuous object consisting of two distinct objects, Bill and Biff. But, according to the objection from unavailable representation, none of the interpretations in (3)–(5) can make (2) come out true. If (3) is true, then how does *A*

[7] See, e.g., Lawlor (2001); Millikan (2000); Perry (2012); Recanati (2012). It is not to my purpose to evaluate the theory of mental files as such here. The framework is part and parcel of the objection, as I understand it, and it is assumed here for sake of argument.

manage to think a thought 'of Biff'? As Camp (2002: 33) would put it, if (3) is true then *A* 'cannot think anything at all' of Biff and, therefore, cannot believe *that Bill = Biff*. *A* could only represent the belief *that Phil*$_c$ *= Phil*$_c$, but, assuming (3) is true, that is clearly not the same as believing *that Bill = Biff*. Same applies *mutatis mutandis* when (4) is assumed.

But what about (5)? Well, according to Millikan this is similarly problematic. For (2) to be true on this assumption, *A* 'should have to have a thought of Bill and another of Biff, which thoughts [*A*] was disposed to coidentify. But a thought of Bill that is other than [*A*'s] thought of Biff is exactly what [*A* does not] have' (Millikan 1994: 97). Therefore, again, (2) cannot be true and the Frege model is demonstrably false.

Camp and Millikan both conclude that combinatory confusion cannot consist in having a false identity belief. But ultimately the objection does not work. There are two reasonable responses to it. First, one could insist that the singular term 'Phil$_c$' is ambiguous. There will be possible contexts in which the speaker definitely intends to refer to Bill by an utterance containing 'Phil$_c$' and contexts where Biff is clearly intended. This seems possible because the contextual salience of the other person can be so low as to make the confusion practically irrelevant (more on this in §4.1 below). Thus some tokenings, on some occasions, may be about Bill and some tokenings about Biff (and perhaps some are also about the Bill/Biff amalgam). The ambiguity is, surely, lost on *A*, but others might become aware that in *A*'s c-idiolect 'Phil$_c$' sometimes refers to Bill and sometimes to Biff. Then it is, at the very least, possible for *A* to believe falsely, and explicitly, *that Bill = Biff*.[8]

Neither Camp nor Millikan takes this option at all seriously. But this is of course a feature of the popular mental files model of singular cognition: each individual object is causally connected to a singular file or concept—and any confusion leads to a kind of corruption. The linguistic item in question is tied to that singular mental file. Still, once the distinction between c-languages and t-languages has been made clear it seems perfectly reasonable to suppose that combinatorily confused singular terms are ambiguous since they are always paired with (at least) two unconfused terms in the t-language (more on this presently). Yet, to construct a stronger argument, I will assume that 'Phil$_c$' cannot be ambiguous.[9]

The concession poses no problem, however. The objection from unavailable repre-

[8] To illustrate, consider the following example. 'Bank' is ambiguous. Assume that *A* falsely believes 'bank' is not ambiguous and refers only to financial institutions. *B* says to *A*: 'Meet me at the bank,' intending to refer to a river bank. *A* may then explicitly form the following false belief: *the concept of 'bank' employed by B is identical to the concept of 'bank' employed by me.*

[9] Indeed, if 'Phil$_c$' is an item in some mental representational system then, or so some have argued, it cannot be ambiguous in the way described. Thoughts do not appear to be ambiguous in the way natural language sentences so appear (see, e.g., Fodor 1978: 198–200).

sentation can be countered more powerfully without making any assumptions at all about the actual content of a confused singular term like 'Phil$_c$.' The question, What is the actual referent of confused singular terms? is not as fundamental as Camp and Millikan seem to think. Confused identity can be captured and defined without committing to a specific view on that thorny issue. Armed with the Frege model and the distinction between c-language and t-language, this is exactly what I propose to do.

Let us start by making the distinction itself more precise. This can be done by introducing two methodological constraints into the dialectic. First, suppose that for any confused singular term a_c in the c-language, a_c can only be mentioned and never used in the theorists' t-language. Otherwise, we will automatically assume that the theorist is confused. That is clearly to be avoided. Second, for any confused singular term a_c in the c-language, the t-language must contain unconfused counterparts. An unconfused counterpart is a singular term a_t in the idiolect of an unconfused speaker, such that if the confused speaker became privy to the truth, they would probably, from then on, intend to use a_t to refer to the object thereby referred to by the unconfused speaker. So when I explain that Bill$_t$ and Biff$_t$ are different people, the confused hearer should try, from then on, to incorporate 'Bill$_t$' and 'Biff$_t$' into their idiolect while banishing the use of the corresponding 'Phil$_c$.' Generally speaking then, in combinatory confusion, c-term a_c is paired with two t-terms, a_t and b_t: $\langle a_c, \langle a_t, b_t\rangle\rangle$. Normally, in separatory confusion, two c-terms, a_c and b_c, are paired with two unconfused t-terms, a_t and b_t, thus: $\langle\langle a_c, b_c\rangle, \langle a_t, b_t\rangle\rangle$. (Think of the 'Cicero'/'Tully' example here.)

If such expression-pairs occur in the wild, *all* of them may turn out to be homophones. In Kripke's (1979: 153) example of separatory confusion, the c-speaker mistakes Paderewski for two distinct individuals, calling each 'Paderewski.' This c-language fragment then contains two confused names 'Paderewski$_{c1}$' and 'Paderewski$_{c2}$' paired with only a single unconfused counterpart—a special feature of homophonic separatory confusions—'Paderewski$_t$' in the t-language. The inverse Paderewski case (cf. Recanati 2012: 141–142) is a case of combinatory confusion involving homophones. Imagine another c-speaker who takes two individuals, P1 and P2, to be one and both happen to be named 'Paderewski.' Here we must assume, as in the Bill/Biff case, that neither P1 nor P2 can lay a claim to being the 'dominant causal source' of the c-speaker's practice of uttering the name or of the attendant mental representation (cf. Lawlor 2007: 162). That is, P1 and P2 are equally responsible for causing the relevant representations of the c-speaker. In this case the c-language contains a single name, 'Paderewski$_c$,' which is paired with two unconfused counterparts, 'Paderewski$_{t1}$' and 'Paderewski$_{t2}$,' in the t-language. I will try to avoid homophonic examples in what follows, for sake of sanity and all that is good and holy.

This way of making the distinction between t-languages and c-languages is general enough to be accepted by both sides of the debate. On the Frege model, however, the

relation between the two languages runs deeper. It is committed to the claim that the t-language must be used to accurately describe the mental state of a thinker who is using a c-language as a medium of conscious thought or a means of communication. So, for example, if there is a language of thought, some individual formulas in the confused speaker's language of thought are best described in a t-language. Confused speakers systematically misrepresent their internal representational systems. But, of course, this is just what the Millikan model denies. On that view, the internal representational system is best described via the speaker's own c-language idiolect. Notice, however, that neither model depends on the language of thought hypothesis being true. If it is false, the models can still claim that they are the best theoretical descriptions we have, at a certain level of abstraction, of the underlying nature of the relevant mental states.

Now the argument from unavailable representation can be properly refuted. In cases of confusion the relevant singular representations of the agent must be bracketed as scrambled and corrupt by the theorist. In particular, such c-terms cannot be incorporated directly into the t-language (of course they can always be mentioned and put into quotes). But as theorists, we want to give the clearest explanation of the agent's behavior, communicative or otherwise. We also want to be able to say what is happening when, as we would put it in the t-language, the confused person *wonders* whether, or *supposes* that, two objects are the same or one object is actually two. And, unsurprisingly, attributing false identity beliefs to agents who may themselves be quite unable to represent those beliefs explicitly will do the explanatory work required. Thus, when *A* combinatorily confuses Bill_t and Biff_t, as shown in dispositions to make certain inferences and assumptions in particular contexts, the theorist ought to explain *A*'s behavior by attributing to *A* an *implicit* identity belief that happens to be false. For instance, *A* will have a complex variety of dispositions to think and express false identities where the identity sign is flanked by two demonstratives 'this' and 'that.' One of them may refer to an object of current perception while the other refers to an object represented by a memory image. Further, when *A* wonders whether Bill_t is Biff_t we can neatly describe *A*'s mental state in the t-language as: wondering whether Bill_t and Biff_t are one or two people.

One can easily allow, then, that the confused agent could indeed have thoughts that are 'of Bill' and thoughts that are of 'of Biff.' Also, remember that the agent's 'Phil_c'-representations have two unconfused counterparts in the t-language, 'Bill_t' and 'Biff_t.' And these representations are the perfect fit for the theorist who wishes to explain *A*'s 'Phil_c'-related thought and behavior. Echoing what Stuart Hampshire said in a similar context, someone may be unable to explicitly think that *p*, while their behavior can only be explained by the hypothesis that they believe that *p*, given that it is known

that they believe that *q*.[10]

If one *explicitly* believes *that p* at time *t* then, let us suppose, some corresponding mental representation is tokened in one's internal 'belief box' at *t*. The mental representation, further, has a content that is specified by the *that*-clause (cf. Fodor 1987). Philosophers recognize at least two kinds of *implicit* belief. First, there are *derived* beliefs, namely beliefs that can be swiftly derived from the set of one's explicit beliefs. Thus, if one believes explicitly that the number of planets is 8, one thereby believes implicitly that the number of planets is lower than 50 (cf. Schwitzgebel 2013). This characterizes the relevant beliefs of the logician Lois #2 discussed in §2.1 above. Secondly, there are *governing* or *guiding* beliefs or attitudes. These can be rules, laws, or biases in accordance with which mental processes move from one explicit representation to another (e.g. Dennett 1978: ch. 6; Fodor 1985). Confused beliefs are more like the latter kind. Just like many types of bias they can be implicit and unconscious, but they can still become the objects of explicit belief later. The agent is, however, unable to swiftly derive the confused belief from their set of explicit beliefs. But postulating the belief is the best way we have, as theorists, to describe their relevant mental states and explain and predict some of their actions. Consider a very simple example. Suppose Bella confuses Bill_t and Biff_t and uses 'Bill_c' for both of them. She meets Biff_t and says, 'Hi Bill_c.' How do the folk explain why Bella uttered the wrong name? Well, simply by saying that she believes that *that man in front of her*, i.e. Biff_t, is identical to Bill_t. People give such explanations all the time, and this practice does not appear suspect.

The Camp-Millikan argument could only be evidence for the view that confused speakers can have no *explicit* beliefs of the form 'X is Y,' but according to the Frege model the belief in question can be either explicit or implicit. The point can be made by saying that, since the theorist should not be allowed to simply adopt the language of the speaker in question, one is compelled to endorse a kind of local interpretationism about confused beliefs. Global Dennettian interpretationism holds, roughly, that a creature's belief *that p* is constituted by the fact that attributing this belief to the creature explains its behavior in an efficacious and simple manner. This, it is hoped, can be done without committing to substantive theses about an internal representational system or language of thought (cf. Dennett 1982, 1991).

The local interpretationism envisaged here, as applied to object-confusion, holds that the Frege model provides the best and most intuitive theoretical description of the confused thinker's internal mental state, at a particular level of abstraction. The view

[10] Cited in Dennett (1982: 164n16). Anticipating the problem of this paper, Hampshire goes on to write in the sentence following: 'Perhaps the confusion in his mind cannot be conveyed by any simple account of what he believes: perhaps only a reproduction of the complexity and confusion will be accurate' (1975: 123).

is thus at home with representationalist theories of mind and should be construed as a theory of the nature and structure of confused mental states. Interpretationism enters the picture at the level of public language, both when it functions as a medium of conscious thought and as a means of communication. The Fregean insists that these linguistic items need to undergo reinterpretation, for they do not accurately reflect the underlying mental states. Specifically, the theorists needs the resources of the Frege model to plausibly describe the mental state of a speaker who utters a confused singular term or employs it to think confused thoughts, even if it is in terms which are not directly available to the thinker. Generally speaking, there is nothing particularly special about this, since the thinker's own construal of their mental states should not always be privileged over the theorist's anyway (more on this in §3.1 below).

Global interpretationism need not be embraced here, as other unconfused singular terms will be shared between the speaker and the theorist. And the theorist can, then, posit explicit internal representations and mental states without any reinterpretation of the speaker's public language expressions. In that case, the theorist does not automatically commit to any errors by using the same public language singular terms as the one who is confused.

It is quite natural to think that the objection from unavailable representation could not be made in terms of separatory confusion. I suspect that this is the view implicit in Camp and Millikan. As a reviewer for this journal notes,

> [w]ith separatory confusion, there *are* enough concepts to allow for the formation of intelligible identity statements, and so the Frege model can stand here, even without invoking the resources of implicit belief, at least so far as the objection from unavailable representation goes.

Clearly, this would make the objection to the Frege model even weaker but, by my lights, this natural thought is not quite true. Explaining how this is so helps to clarify the difference between the two models of confusion. In combinatory confusion, according to Millikan, the confused thinker has no conception of an object which is not also, at the same time, a conception of some other object. This makes the individual concept itself corrupt and confused. By parity of reasoning, in separatory confusion, the confused thinker should have no conception of an object which is not also, at the same time, a conception of something which is *not* that very object. So, Lois Lane has no conception of Superman, the thought would go, which is not also, at the same time, a conception of something which is *not* Clark Kent, i.e. *not* Superman. This would, according my construal of the Millikan model, make the concept confused. It should be clear, however, that this would complicate the dialectic considerably, and neither Millikan nor Camp seems to endorse an argument of this sort, thus I focus

on combinatory cases. Millikan's broader theoretical commitments may even make such an argument difficult to formulate within a teleosemantic framework. Perhaps the complications are such as to make the objection from unavailable representation less plausible overall but, even so, this would clearly not undermine my overarching point in this section.

3.1 Descriptive and explanatory adequacy

As emphasized by Ruth Marcus (1983), if a theory of confused belief is to be descriptively adequate it must make room for the fact that nonlinguistic animals and prelinguistic infants are just as confusion-susceptible as other more intellectually sophisticated creatures. Despite appearances to the contrary, the Millikan model holds no advantage in this respect. Using the t-language, the Fregean theorist can assign implicit identity beliefs to nonlinguistic agents if such are required, for example, to explain their behavior. Camp, Marcus, and Millikan seem to be driven, in different ways, by the intuition that confusion-susceptibility is cognitively more basic than the ability to think thoughts about or involving identity. If thoughts can be implicit, either in the sense of being *derivative* or *governing* propositional attitudes, the intuition is arguably mistaken. On the Frege model the mental state of confusion is explained in terms of such attitudes even when the subject is not in explicit possession of the concepts of identity and distinctness of objects.

Strictly speaking, the Frege model and the Millikan model, as presented here, are only intended as theories of the mental state of *confused* identity, i.e. a certain type of cognitive malfunction. Obviously, however, the two models will mesh differently with theories of corresponding *nonconfused* mental states, i.e. theories explaining our capacity to successfully reidentify objects in thought or perception. Indeed, according to Millikan herself the '... central job of cognition is the ... task of reidentifying individuals, properties, kinds, and so forth, through diverse media and under diverse conditions' (2000: xi). And her theory of confusion is certainly intended to complement her theory of this central cognitive ability.

In this section I will briefly sketch how one might naturally extend the Frege model to cover successful reidentification and contrast this with what I take to be Millikan's account. There is not enough space here to argue that the Frege model is right and the Millikan model is wrong about these cases and, fortunately, this is not necessary for the purposes of this paper. The proponent of the Frege model can in fact embrace much of the Millikan model, only insisting, along the way, that implicit or explicit belief in propositions involving identity ought to be postulated when possible and explanatorily fruitful. We have already seen how this works in cases of confusion but the Fregean will, at the very least, maintain that some cases of successful reidentification are to be

explained by attributing propositional attitudes as well.

According to Millikan's more functionalist picture, reidentification is explained by the capacity to recognize when two thought tokens are thoughts of the same. And, surely, this does not imply that the agent must explicitly represent *that the subject of one thought token is identical to the subject of another thought token.*[11] I note here that part of her argument for this view is, precisely, that it is impossible to describe confusion in terms of belief states. As we have seen this is wrong.

The Fregean, by contrast, will either hold that any case of successful reidentification is best explained by appeal to the assumption that the thinker implicitly believes in a true proposition involving identity, or that at least some cases must be so explained. I illustrate the stronger version here—noting where it might be weakened.

To borrow Fodor's (1985: 24) example, let us assume some type of associationism about mental processes. Imagine, further, that there is a 'principle of association by proximity' in virtue of which thoughts of salt are usually associated with thoughts of pepper. The principle is then a guiding or law-like attitude which explains why salt-thoughts and pepper-thoughts are invariably linked in trains of thought. But the principle itself need not be explicitly represented in the mind. According to standard representational theories of the mind, however, the condiment-thoughts themselves must be explicitly represented.

Something similar happens, on the Frege model, whenever a train of thought contains different thoughts that are supposed to be about the very same object. This can occur, for example, in inferences. When a thinker infers *that Cicero is an orator* from the belief *that Cicero is tall and an orator*, they assume that they are thinking about the same object twice. Although there is no need to assume that they explicitly believe or entertain the thought *that Cicero = Cicero* it helps to explain their inferential behavior if we assume that they believe this implicitly. Here, the weaker variety would hold that such an implicit belief is only needed when the inference involves two names which the thinker takes to be distinct. So, we would need the belief *that Cicero = Tully*, for example, to explain success in such a case. And if—as Millikan asserts in one of her discussions (1994: 97)—we are not allowed to attribute any such belief to them, explicit or implicit, it would be unclear how to explain this piece of behavior at all, other than referring to it as the manifestation of a 'capacity.'

On the stronger extension of the Frege model, implicit identity beliefs, when true, always form part of the explanation of simple inferences like the one from *Fa* to *Fa*.

[11] As John Campbell (2002: 97–101) likes to put it, when one identifies o at t and reidentifies o at a later time t' one must at least be 'trading on' the identity of o at t and t'. Relatedly, Tyler Burge (2010, e.g. 286–287, 460) argues extensively that creatures need not be capable of *thinking* the 'criteria for reidentification' in order to reidentify individuals and objects. Mere 'perceptual tracking' counts as reidentification.

And, correspondingly, such inferences can fail on account of the thinker's *false* identity beliefs regarding *a* or, if you will, on account of the thinker's *lack* of the true beliefs. It is a virtue of the stronger version that it gives a uniform account of success and failure characteristic of mental representations involving identity and distinctness. In separatory confusion the thinker *fails* to infer *Fb* from *Fa* and the unconfused succeeds in *doing* so. So Lois Lane fails to infer facts about Clark Kent from facts about Superman. In combinatory confusion the thinker *fails* by wrongly inferring *Fb* from *Fa* and the unconfused successfully *blocks* any such inference. So *A* wrongly infers facts about Biff from facts about Bill (as we describe it in the t-language), i.e. wrongly trades on the identity of non-identical objects.

As already mentioned, I will not argue for this extension of the model here, and there are many strong and intuitive objections to consider. For example, one might suppose that explaining reidentification in this way leads to an infinite regress since one would always need a *further* belief in identity to explain the move from a term in the inference to the terms in the identity belief itself, and again *ad infinitum*. Furthermore, it may seem like when a thinker successfully 'trades on' identity in very simple inferences, postulating the identity belief would be explanatorily superfluous. I hope to address these worries elsewhere, but here I only observe that the weaker version of the Frege model is not subject to them and may represent a middle ground between the two views.[12]

Still, I want to develop two general considerations suggesting that one ought to prefer the Frege model as a theory of *confusion*. Firstly, those, like Millikan, who suppose that accepting or rejecting identities reduces without remainder to acts of connecting or disconnecting dots or files in a mental network tend to think that there is no such thing as judgments of identity. Thus, she writes that,

> as distinguished from an identity statement or assertion, there is no such thing as an identity *judgement*. It is not the job of an identity sentence to induce a belief. Its job is to induce an act of coidentifying. (2000: 172)

As Papineau and Shea (2002: 463) aptly point out, after quoting this very passage, we bear many attitudes to propositional identities other than acceptance and rejection. For example we can wonder whether Bill is Biff or entertain the supposition that Bill is Biff, and these propositional attitudes do not seem to reduce neatly to acts of connecting or disconnecting mental files (or dispositions to perform such acts). Propositional identities do have, as Millikan argues, distinctive effects on the functional organization of mental states. But they can also be the objects of a multitude of propositional attitudes. Thus there is reason to doubt that Millikan's view has the resources to

[12] I want to thank a reviewer for this journal for helpful comments on these issues.

explain all the psychological phenomena associated with identity beliefs or identity statements.

No such worry attaches to the Frege model. There are identity beliefs, even if they must sometimes be couched in terms different from those employed by the believer, if the believer has a language at all. But this is true more generally: beliefs of pre-linguistic humans and nonhuman animals being among the clearest examples. Local interpretationism holds that the same can be true of rational adults. Their utterances need to be systematically reinterpreted in order correctly describe the underlying mental state. In a forthcoming paper, I argue that many cases of malapropism should be treated in like manner (Unnsteinsson forthcoming). I call this the misarticulation theory of malaprops. For example, when John Kerry slips and utters, 'Wasabi is a dangerous sect,' but meant to utter the word 'Wahabi' instead of 'Wasabi,' we ought to construe him as actually having expressed the proposition he intended to express and not the one he apparently expressed. /wasabi/ was, on this occasion, his erroneous way of pronouncing /wahabi/. This is true even if the slip is persistent, i.e. one which occurs consistently and repeatedly in Kerry's speeches. Object-confusion is a more deep-seated error but it also requires some degree of reinterpretation: to describe the mental state of the speaker adequately one must avoid taking the words uttered at face value.

Thus the Frege model surpasses the Millikan model in explanatory power, if indeed the latter eschews implicit identity beliefs altogether. And this brings us to the second consideration. It seems reasonable to think that no model of confusion can genuinely avoid postulating identity beliefs or some roughly equivalent propositional attitude. The reason for this is the distinction between t-languages and c-languages. If the theorist is to remain unconfused while explaining the behavior, inferential or otherwise, of a confused agent, they can only ever *mention* confused terms from the c-language. But if this requirement is accepted, and I see no reason to why it should not, the Millikan model must postulate false identity beliefs or similar belief-like states. To see this, let us try to state the Millikan model as precisely as possible.

Millikan model of confusion

Speaker A suffers from identity confusion at time t iff for any propositional attitudes V and W that A explicitly holds at t, either (1) or (2) is true:

1. (i) A *Vs* F(the referent of 'a_c' in A's c-idiolect), (ii) A *Ws* G(the referent of 'b_c' in A's c-idiolect), and (iii) $a_t = b_t$ but A's 'a_c'/'b_c'-representations presuppose that $a_t \neq b_t$.
2. (i) A *Vs* F(the referent of 'a_c' in A's c-idiolect), and (ii) $a_t \neq b_t$ but A's 'a_c'-representations presuppose that $a_t = b_t$.

In (1) the speaker suffers from separatory confusion and 'a_c' and 'b_c' are paired with corresponding terms in the t-language. For instance, 'Superman' in the c-language is paired with 'Superman' in the t-language, and so on. (2) is combinatory confusion and 'a_c' is paired with two corresponding terms in the t-language. If A uses 'Bill_c' to refer to what the t-speaker refers with both 'Bill_t' and 'Biff_t' these are paired together. Once the distinction between the c-language and t-language is made there is no reason to keep to the terminology that was introduced before in explaining the Millikan model. That is, we need not say, in (1), that the c-speaker refers to a spatially discontinuous object consisting of two distinct objects. Nor that, in (2), the c-speaker purports to refer to a nonexistent object that consists of itself and not itself at the same time. Millikan's model aimed at bringing in the actual content of items in the confused speaker's inner representational system. But this is not necessary, as the mental state of confusion can be defined without mentioning those contents. This is good since the model need not take a stand on particular content-assignments in these puzzling cases.

But what does it mean to say that representations in the c-language 'presuppose that $a_t = b_t$ or that $a_t \neq b_t$'? Roughly, that the way in which A uses these representations is best explained by assuming that A implicitly believes one or the other. The representations 'function as if' they were representing a single object while they actually represent two objects. Or the representations 'function as if' they were representing two objects while they actually represent no object at all. In other words, the relevant parts of A's behavior are neatly explained by assuming that their explicit mental representations are governed by false identity claims. Since theorists are not restricted to using the language of the subject under discussion they are free to use the identity sign to capture these facts in a clear and precise fashion. And nothing seems to compete, in terms of clarity, with using the identity sign in these cases. The Millikan model tries to describe confused identity merely in terms of (corrupt) functional organization of mental states, rather than in terms of propositional beliefs. Remember, however, that the argument from unavailable representation, on which the model is based, only purported to show that identity beliefs were in some sense impossible in cases of confusion. Since this argument fails the motivation for doing away with propositional belief is no longer a factor and we are free to use them to explain corepresentational capacities and their characteristic failures.

4 Proper functions and malfunctions

Given that I've shown that the Frege model best explains the mental state of confusion, we don't need to construe confusion as 'an error of its own kind' (Millikan 2000: 173), as opposed to a false belief. Even so, I'll now go on to argue that confusion in fact

constitutes a distinctive kind of malfunction, one that we can even characterize in something like Millikan's own terms while still accepting the Frege model as the true metaphysics of confusion.

First: what is a proper function? The notion is, of course, borrowed from evolutionary biology (Millikan 1984, 1989b, 1989a). One item can serve many different functions at the same time but usually only a subset of these functions actually helps explain why the item continues to be reproduced. Consider the human heart. It seems to have many different effects on the body and environment. For example, the heart makes the human body heavier, emits a low thumping sound, and pumps blood. Only the last is plausibly thought of as the proper function of human hearts, because it is historically responsible for the fact that hearts are reproduced. Importantly, however, an item that does not actually pump blood can still be a heart. Malformed and malfunctioning hearts are still 'supposed to' pump blood, i.e. serve the proper function whereby their proliferation is evolutionarily and historically explained. Going further, Millikan applies the notion of proper function to biological and cultural items alike. Thus, according to her, linguistic devices (words, syntactic forms, etc.) have their own proper functions. For example, the proper function of the indicative mood is the production of a belief in the hearer that corresponds to the meaning of the indicative sentence uttered. The imperative mood has the proper function of producing compliant behavior. If I say 'Pass the salt' and my interlocutor then passes the salt, my imperative utterance has successfully performed its proper function. The proper function can also be called the 'stabilizing' function of the item, since it is responsible for perpetuating the occurrence of the item in a linguistic community.

Gloria Origgi and Dan Sperber (2000) argue that Millikan's theory gets this part wrong because it is not necessary that imperative utterances typically or reliably produce compliant behavior, or even that they produce in the hearer a desire to comply. On their view it is more plausible to say that imperative utterances function properly when the speaker succeeds in guiding the hearer towards the correct interpretation, so the hearer understands which course of action would satisfy the imperative utterance. Compliance is an additional step that depends entirely on the hearer's own beliefs, desires, and intentions.

But even if imperatives produced compliance only on very rare occasions, it might still be the case, at least on Millikan's theory as I understand it, that their proper function is to produce compliant behavior. Millikan has emphasized that proper functions have nothing to do with typicality, reliability, or statistical averages (1989b: 21–22; 1989a: 93–94; 1984: 4, respectively). All that is required, on her view, is that an explanation of why imperatives are reproduced essentially invokes historical occasions on which imperatives in fact caused compliance. A sequence of such occasions would then explain why imperatives were selected for by a process analogous to natural

selection. Millikan defines 'normal conditions' as the actual conditions that have been historically needed for an item to perform its proper function. In the case at hand, the normal conditions might involve things like the hearer's understanding, the speaker's command of the hearer's language, and the hearer's willingness to do as told.[13] The fact that hearts function to pump blood is explained by a history of normal hearts pumping blood in the normal way under normal conditions. As Millikan (1984: 56) puts it: 'If no token of the imperative mood ever effected more than an abortive attempt or intention to comply with it, it is clear that speakers would soon cease to use the imperative forms at all or to use them as they now do.'

On Millikan's (2000) account the distorting effects of confusion on basic cognitive processes are fairly straightforward. As mentioned above, she argues that a central task of cognition is to reidentify particulars and properties in thought and perception. Thus, on her view, we can postulate a basic cognitive mechanism in the mind/brain with the proper function of reidentification. The mechanism itself, let us suppose, has proliferated and is perpetuated in humans by a process of differential reproduction. The mechanism underlies many important cognitive tasks, such as recognition, expectation, and inference (cf. Lawlor 2001; Recanati 2012).

If we can describe such a mental mechanism, at some requisite level of abstraction, combinatory and separatory confusions—as defined by the Frege model—are clearly the most characteristic ways in which it breaks down and fails to perform its proper function. Its function is to keep track of particulars in the world and false identity beliefs give rise to systematic and characteristic patterns of cognitive failure. For example, if *A* expects Bill (or Biff, or Bill/Biff, or the referent of 'Bill' in *A*'s language, etc.) to knock on the door in five minutes while implicitly believing falsely *that Bill = Biff*, then *A*'s reidentificatory capacity is disrupted because *A*'s belief makes the relevant identifications unreliable. Repeated and chronic confusion indicates that the underlying mechanism itself is impaired.

4.1 The proper function of singular terms in communication

Finally, let us apply the notion of proper function to singular terms in a natural language. Millikan holds that the stabilizing function of a proper name—a paradigm example of a singular term—is to 'precipitate an act of identification of its referent' (1984: 80). At a high level of abstraction this sounds plausible. But I much prefer to describe the function of names in the more Gricean terms of utterances providing evidence for the speaker's referential intention.[14] But, to keep the argument simple

[13] Cf. Godfrey-Smith's (1994: 265–266) discussion of the neck-expanding display of the frill-necked lizard.

[14] Cf. Bach (1987); Grice (1989); Schiffer (1981); Unnsteinsson (2014, 2016); Wilson & Sperber (2012).

and intuitive I have stated it in Millikan's own terminology, only reverting to the more detailed Gricean vocabulary when needed. In my view, the theories complement each other quite naturally.

Intuitively, if all singular term tokens would become confused, speakers would soon cease to utter them or they would acquire some distinct function. This intuition seems, however, to depend on the assumption that members of the linguistic community in question are all confused *in different ways about different things* or that some are confused and some are not. Matters are very different when a whole group of speakers, for example a whole scientific community, is uniformly confused about the identity of objects or properties to which they intend to refer in speech (Camp 2002: ch. 2; Evans 1982: ch. 11; Field 1973). As Hartry Field argues, using 'mass' as an example, such global confusion of two different properties or natural kinds makes the reference of a term metaphysically indeterminate. Although such cases are interesting in their own right, I want to leave them to one side. I want to argue, rather, that in cases where the confused agents are perfectly capable—by being a bit more careful or discerning—of seeing things aright, the function of singular term utterances becomes corrupt in principle. More precisely, the utterance will precipitate two conflicting acts of identification (combinatory confusion) or no such act at all (separatory confusion). And this is not due to Fieldian indeterminacy.

Thus, to focus on local rather than global confusions, assume that the speaker's (A) language is a c-language and the hearer's (B) language is an unconfused t-language. First, I give the argument in terms of combinatory confusion. Then we move on to separatory confusion.

Corruption in combinatory confusion. Assume, first, that A utters '. . .Bill_c . . .' on a particular occasion. And, secondly, that A believes falsely *that* $\mathit{Bill}_t = \mathit{Biff}_t$. In that case, A's utterance simultaneously precipitates the identification of two distinct objects while its normal stabilizing function is to precipitate the identification of a single object. And this is precisely because of A's false identity belief: if A is thoroughly confused, both individuals will equally count as what A intended to refer to. Someone may object by saying that, actually, the utterance only precipitates the identification of Bill_c, i.e. the man named 'Bill' in A's c-language, whoever that is. But this is wrong, especially if beliefs determine or constrain referential intentions. To see this consider a normal unconfused case. I know the truth about Bill_t not being Biff_t. If I utter '. . .Bill. . .,' intending to refer to Bill, who is not identical to Biff, my utterance does not precipitate an act of identification of Biff at the same time. Since my belief is true, no problem seems to arise. In a similar way, in virtue of A's false identity belief, A's utterance precipitates the identification of two *distinct* objects, corrupting the proper function of the utterance.

This argument naturally suggests the following question. What happens when B is

fully aware of the extent of *A*'s confusion? Doesn't that get rid of the corruption, since the hearer can presumably grasp what the speaker intends? Well, it is not so simple. Of course the hearer is in a much better epistemic position to discern the speaker's mental state, but clearly the function of the speaker's singular term utterance becomes corrupted in the the same manner as before. There are still two objects in contention for identification, since the hearer also knows—remember that we are not considering global confusion here—that the speaker believes *falsely*. And if Millikan and others are right about the proper function of singular terms only a single object should be in contention.

Should we say instead, perhaps with Millikan on our side, that *A*'s utterance of '. . .Bill_{c} . . .' in fact functions to precipitate the identification of the $\text{Bill}_{\text{t}}/\text{Biff}_{\text{t}}$ amalgam? At least this would be a single object of some sort. I think not. It helps to use Gricean terminology to explain this point clearly. Uttering an ordinary proper name of English, for example, would never constitute good evidence for an intention to refer to a gerrymandered object of this kind. Unless, perhaps, both speaker and hearer are confused in the exact same manner (more on that below). If I alone have the crazy belief that the apple in my hand is really a mereological fusion, *o*, of the apple and the Empire State building, I cannot expect my audience to get the reference to *o* by simply uttering 'This apple tastes great!' And if it is true, as I think it is, that communicative intentions are severely constrained by doxastic states, then I cannot even intend for my audience to get the reference to *o* since it is quite likely that I will believe, in such a context, that it is impossible that they will comprehend me as referring to that gerrymandered object with that expression (see, e.g., Donnellan 1968).

Corruption in separatory confusion. Using the fictive world of Superman as an example, assume now that *A* utters '. . .$\text{Superman}_{\text{c}}$ . . .' on a particular occasion. And, secondly, that *A* believes falsely *that* $\textit{Superman}_t \neq \textit{Clark Kent}_t$. In that case, *A*'s singular term utterance precipitates the identification of an object *o* while, at the same time, precipitating the identification of an *x* such that $x \neq o$. Specifically, there is no object which is Superman while not being Clark Kent and, so, the utterance precipitates the identification of no actual object while its normal stabilizing function is to precipitate the identification of a single object. And this is precisely because of *A*'s false identity belief. As before, to see this, it helps to consider a normal unconfused case. I know the truth about Superman being Clark Kent. If I utter '. . .Superman. . .,' intending to refer to Superman, who is identical to Clark Kent, my utterance also precipitates an act of identification of Clark Kent. Since my belief is true, no problem seems to arise.

One objection in particular may come to mind here. And the answer is made easier by using the more detailed Gricean terminology. Is it really the case that *A*'s utterance, in the above context, is evidence that *A* refers or intends to refer to an object *o* such that $o = \textit{Superman}$ *and* $o \neq \textit{Clark Kent}$? Yes, it seems so. For the following is a generally

sound principle: if E is evidence that A intends to refer to a, and it is true *that* $a = b$, then E is evidence that A intends to refer to b. Bear in mind that the notion of *intending to refer to a* should be construed transparently. The principle would *not* hold water if it were substituted for an unequivocally opaque notion, such as *intending to refer to a as a* or *intending to refer to a as b*. Similarly, while *seeing* is transparent, *seeing as* is opaque. If I see John and John is my cousin, I also see my cousin. But seeing John as John does not imply seeing John as my cousin. By the same token, if I intend to refer to John and John is my cousin, I also intend to refer to my cousin. This remains true even if I utter (confused) sentences like 'My cousin and John are not the same person.' But my false beliefs may, as in the two cases above, create a conflict in my referential intention and corrupt the proper function of the singular term.

Someone like Field may object that my whole argument has a hidden false assumption, namely, that the locally confused speaker can have a determinate referential intention at all. Local confusion is just like global confusion: it gives rise to metaphysical indeterminacy. Field, I imagine, would go on to argue that there is no point in talking about *conflicting evidence* for an *indeterminate* conclusion. On this account, then, the combinatorily confused speaker, for example, has an intention that refers indeterminately to Bill_t, Biff_t, and the $\text{Bill}_\text{t}/\text{Biff}_\text{t}$ amalgam, and an utterance involving 'Bill_c' or 'Biff_c' (or, as in our original example, 'Phil_c') is good evidence for this indeterminate intention.

The objection misses the point of the present exercise. If we consider ordinary proper names and perceptual demonstratives, rather than natural kind predicates like 'mass' in Newton's language, it is clear that even the confused speaker *can*, on occasion, have a very determinate intention to refer to an object about which they are confused. But, in those cases, the part of the linguistic evidence that consists in an utterance of a confused token of a singular term will still be corrupt, just with a diminished practical upshot. Here is the kind of case I have in mind. A confuses Bill_t with Biff_t but stands right in front of Bill_t and utters 'Bill_c is right there' while pointing directly at him. Assume also that Biff_t's contextual salience is quite low. In this case, I suppose, A could definitely intend to refer to *that man there*, i.e. Bill_t. Without doubt, A also intends to refer to Biff_t, but this is such an insignificant part of A's communicative intention that it seems well nigh irrelevant.

To summarize, the mere presence of false identity beliefs can make it such that singular term utterances cannot perform their stabilizing proper function of precipitating an act of object-identification. It is reasonable to suppose, then, that the reproduction of such terms in natural language needs to be explained by invoking a sequence of utterances where such confusion was not present at all.

5 Conclusion

I have argued that identity confusions can be defined as a mental state, characterized by certain propositional attitudes. Minimally, the confused agent must either not believe a true proposition about identity or, less minimally, believe a false proposition about identity. The most serious objections to this account can be staved off by arguing that the beliefs can be implicit, but still, the agent must stand in some explicit attitude relations to the object(s) in question. Furthermore, I showed how this mental state has disrupting effects on the proper function of singular terms in communication.

Bibliography

Bach, K., 1987. *Thought and reference.* Clarendon Press, Oxford.

Burge, T., 2010. *Origins of objectivity.* Clarendon Press, Oxford.

Camp, J. L., 2002. *Confusion: A study in the theory of knowledge.* Harvard University Press, Cambridge, MA.

Campbell, J., 2002. *Reference and consciousness.* Clarendon Press, Oxford.

Dennett, D., 1978. *Brainstorms: Philosophical essays on mind and psychology.* MIT Press, Cambridge, MA.

———, 1982. "Beyond belief." A. Woodfield (ed.), *Thought and object*, Clarendon, pp. 1–95. Repr. in Dennett (1987), pp. 117–202.

———, 1987. *The intentional stance.* MIT Press, Cambridge, MA.

———, 1991. "Real patterns." *The Journal of Philosophy*, 88(1):27–51.

Donnellan, K., 1968. "Putting Humpty Dumpty together again." *The Philosophical Review*, 77(2):203–215. Repr. in Donnellan (2012), pp. 31–48.

———, 2012. *Essays on reference, language, and mind.* Oxford University Press, Oxford.

Evans, G., 1982. *The varieties of reference.* Clarendon Press, Oxford.

Field, H., 1973. "Theory change and the indeterminacy of reference." *The Journal of Philosophy*, 70(14):462–481.

Fodor, J., 1978. "Propositional attitudes." *The Monist*, 61(4):501–523. Repr. in Fodor (1981), pp. 177–203.

———, 1981. *Representations: Philosophical essays on the foundations of cognitive science.* The Harvester Press, Sussex.

———, 1985. "Fodor's guide to mental representation: The intelligent auntie's vademecum." *Mind*, 94(373):76–100. Repr. in Fodor (1990), pp. 3–30.

———, 1987. *Psychosemantics.* MIT Press, Cambridge, MA.

———, 1990. *A theory of content and other essays.* MIT Press, Cambridge, MA.

Godfrey-Smith, P., 1994. "A continuum of semantic optimism." S. Stich & T. Warfield (eds.), *Mental representation*, Basil Blackwell, Oxford, pp. 259–277.

Grice, P., 1989. *Studies in the way of words*. Harvard University Press, Cambridge, MA.

Hampshire, S., 1975. *Freedom of the individual*. Chatto & Windus, London, 2 edn.

Kaplan, D., 1968. "Quantifying in." *Synthese*, 19(1-2):178–214.

———, 2013. "De re belief." R. T. Hull (ed.), *The American Philosophical Association Centennial Series. Volume 9. Presidential addresses of The APA 1981-1990*, The American Philosophical Association, pp. 25–37.

Kripke, S., 1979. "A puzzle about belief." A. Margalit (ed.), *Meaning and use*, Reidel, Dordrecth, pp. 139–183. Repr. in Kripke (2011), pp. 125–161.

———, 2011. *Philosophical troubles: Collected papers, vol. 1*. Oxford University Press, Oxford.

Lawlor, K., 2001. *New thoughts about old things: Cognitive policies as the ground of singular concepts*. Taylor & Francis, New York.

———, 2005. "Confused thought and modes of presentation." *The Philosophical Quarterly*, 55(218):21–36.

———, 2007. "A notional worlds approach to confusion." *Mind & Language*, 22(2):150–172.

Marcus, R. B., 1981. "A proposed solution to a puzzle about belief." *Midwest Studies in Philosophy*, 6:501–510.

———, 1983. "Rationality and believing the impossible." *The Journal of Philosophy*, 80(6):321–338. Repr. in Marcus (1993), pp. 143–161.

———, 1990. "Some revisionary proposals about belief and believing." *Philosophy and Phenomenological Research*, 50 (Supplement):133–153. Repr. in Marcus (1993), pp. 233–255.

———, 1993. *Modalities: Philosophical essays*. Oxford University Press, Oxford.

Millikan, R. G., 1984. *Language, thought, and other biological categories: New foundations for realism*. MIT Press, Cambridge, MA.

———, 1989a. "Biosemantics." *The Journal of Philosophy*, 86(6):281–297. Repr. in Millikan (1993), pp. 83–102.

———, 1989b. "In defense of proper functions." *Philosophy of Science*, 56(2):288–302. Repr. in Millikan (1993), pp. 13–30.

———, 1993. *White queen psychology and other essays for Alice*. MIT Press, Cambridge, MA.

———, 1994. "On unclear and indistinct ideas." *Philosophical Perspectives*, 8:75–100.

———, 1997. "Images of identity: In search of modes of presentation." *Mind*, 106(423):499–519.

———, 2000. *On clear and confused ideas: An essay about substance concepts*. Cambridge University Press, Cambridge.

Origgi, G. & Sperber, D., 2000. "Evolution, communication and the proper function of language." P. Carruthers & A. Chamberlain (eds.), *Evolution and the human mind*, Cambridge University Press, Cambridge, pp. 140–169.

Papineau, D. & Shea, N., 2002. "Ruth Millikan's *On clear and confused ideas*." *Philosophy and Phenomenological Research*, 65(2):453–466.

Perry, J., 2012. *Reference and reflexivity*. CSLI Publications, Stanford. 2nd ed.

Recanati, F., 2012. *Mental files*. Oxford University Press, Oxford.

Richard, M., 2013. "Marcus on belief and belief in the impossible." *Theoria*, 28(3):407–420.

Salmon, N., 1986. *Frege's puzzle*. Ridgeview Publishing Company, California.

Schiffer, S., 1981. "Indexicals and the theory of reference." *Synthese*, 49(1):43–100.

Schwitzgebel, E., 2013. "A dispositional approach to the attitudes: thinking outside of the belief box." N. Nottelmann (ed.), *New essays on belief: Constitution, content and structure*, Palgrave MacMillan, London, pp. 75–99.

Unnsteinsson, E., 2014. "Compositionality and sandbag semantics." *Synthese*, 191(14):3329–3350.

———, 2016. "Wittgenstein as a Gricean intentionalist." *British Journal for the History of Philosophy*, 24(1):155–172.

———, forthcoming. "A Gricean theory of malaprops." *Mind & Language*.

Wilson, D. & Sperber, D., 2012. *Meaning and relevance*. Cambridge University Press, Cambridge.