Rachel Goodman
University of Illinois, Chicago

Confusion and Explanation

(This is the penultimate version of a paper to appear in *Mind and Language*. Please cite the published version.)

The motivating context for Elmar Unnsteinsson's *Talking About* is a challenge to the legitimacy of *intentionalist* theories of reference and meaning: if reference can occur when referential intentions are 'confused' (and would thereby fail to determine the *right* referent, or *any* referent) then how can intentionalism be true? Unnsteinsson's response is, essentially, that confusion *does* in fact interfere with successful reference. To support this claim, he gives an account of the cognitive mechanisms whose proper operation allow *un*confused referential intentions to determine reference, along with a corresponding account of how the failure of these mechanisms, in cases of confusion, disrupts reference. This means that Unnsteinsson's theory of what he calls, 'the mental state of confusion' is central to his project. My aim here is to illustrate what does and doesn't work about this account. To do this, I'll offer an interpretation of a well-known argument from Campbell (1987). The argument has been central in the contemporary literature on Frege's Puzzle and is often glossed in a particular way. However, Unnsteinsson's account of confusion provides an important opportunity for more clarity about how the argument is best interpreted, what it shows, and what is misleading about the way it is commonly presented.

1. Identity Confusion

We should begin with some clarity about our topic. The particular kind of confusion at issue for Unnsteinsson is *identity confusion* and Unnsteinsson counts two kinds of case as species of 'the mental state of [identity] confusion' (Unnsteinsson 2022, Ch. 2).

The first case is what the literature on reference and concepts traditionally calls 'confusion'.[1] This is the case in which, *intuitively* speaking, two (or more than two) things are incorrectly taken by a thinker to be a single thing. For instance, Unnsteinsson works with a case in which he surreptitiously buys his daughter, who wants a teddy bear, two qualitatively identical bears, so that he can switch them out and wash one while she's using the other (Ibid., p. 30). His daughter does not discover that there are two bears, and takes herself to be the owner of a single bear, which she names 'Malcolm'. Despite the fact that her ongoing encounters are sometimes with one bear (we, who know that there are two, can call him 'Bill') and sometimes with the another (we can call him 'Biff'), she does not distinguish them. She goes on to think thoughts of the form, '*Malcolm* sleeps in until I get home from daycare every day', '*Malcolm* wears a blue duffle coat and a red bucket hat', and so forth. Unnsteinsson calls this *combinatory confusion* (Ibid., p. 31).

The second case is what Unnsteinsson calls *separatory confusion* (Ibid., p.31). This is the kind of case with the structure of a 'Frege case': intuitively speaking, a thinker takes one thing to be two (or

[1] See, e.g., Millikan (2000) and Camp (2002).

more than two) things. For instance, famously, Lois Lane takes Clark Kent and Superman to be distinct individuals. Lois's belief to the effect that 'Superman can fly' is in fact about the same individual as her belief that 'Clark Kent works at the Daily Planet', but she does not realise this, and her behavior and rational entitlements reflect this. For example, she is not disposed to infer, and indeed not in a position to *rationally* infer, that one of her co-workers at the Daily Planet can fly.

On the one hand, Unnsteinsson's view that traditional confusion (combinatory confusion) and cognitive Frege cases (separatory confusion) count as species of a single mental state type (*identity confusion*) is natural enough. Intuitively, both kinds of case involve a thinker suffering misapprehensions about the identity of the thing/s she is thinking about. But, Unnsteinsson's motivation to count combinatory and separatory confusion as species of a single mental kind is also theoretical: he wants to argue that the two kinds of case are epistemically or cognitively alike in that they both lead to corruption or malfunction of the speech act of referring (Ibid., p.32 & Ch. 7).

The view that *combinatory confusion* defeats reference is controversial but has strong precedent (Millikan 1994, 2000; Camp 2002). If Elmar's daughter does not distinguish encounters with Bill from encounters with Biff, one might hold there is no *fact* of the matter about which thing her uses of 'Malcolm' (and the thoughts they express) refer to. However, the view that separatory confusion defeats reference is unusual enough to be quirky. That Lois fails to recognize that Clark is Superman does not seem to make it any less determinately the case that her uses of 'Clark' and 'Superman' (and the thoughts they express) refer determinately to the superhero moonlighting as a mild-mannered reporter. So, the commonly held view is that combinatory and separatory confusion have different cognitive repercussions. This in itself might give us reason not to classify the two kinds of case as instances of a single, theoretically important, mental kind.

However, Unnsteinsson's view is that all cases of identity confusion lead to the indeterminacy or failure of reference of the cognitive act of referring. Of course, he realizes the intuitive implausibility of the claim that Lois cannot refer successfully to Superman, and so offers an account according to which her speech acts might pass muster for most ordinary purposes even though they are, in fact, cognitively *defective* in the same way that his daughter's utterances of 'Malcolm' are (Unnsteinsson 2022, p.18-20 & p.158-65).[2]

Giving an account of identity confusion is important because confusion disrupts reference. Giving an account of the success with respect to which confusion is a failure—call it 'clear and unconfused thought'—is important because this success is part of the cognitive mechanism whose proper operation allow referential intentions to determine reference.

2. **The 'Belief Model' of Confusion**

Unnsteinsson defends a 'belief model' of identity confusion, alongside a corresponding belief model of clear and unconfused thought.

The belief model of confusion says that identity confusion consists in a thinker's false beliefs about the identity of the referents of her singular attitudes (or in the lack of relevant and required *true* beliefs about their identity) (Ibid., p.32).

[2] See also Unnsteinsson 2019.

Notice that, in cases of identity confusion, we tend to explain the behavior of a confused thinker by saying things like, 'Lois believes that Superman is not Clark Kent', or 'Elmar's daughter believes that Bill is Biff'. That is, we tend to explain the behavior of thinkers like Lois and Elmar's daughter (including their inferential behavior) by positing false identity beliefs about the things that they are thinking about. The most general idea behind the belief model is that we should take these explanations more or less on face value: they are good explanations, and we can and should use them, not only for ordinary purposes, but to theorise the mental state of confusion (Ibid., p.38). Being confused *consists in* having false, singular identity beliefs about the things you are thinking about.

Though this is the heart of the belief model, the account of course contains more detail. The elaboration of these details is framed as a response to the kinds of worries about the belief model that have led to its being regarded it as a non-starter and have, furthermore, led to fairly wide adoption of what Elmar calls the 'concept model' of confusion.

The first of these worries is that the belief model of confusion is too cognitively demanding. Surely, the worry goes, small children and puppies can take two objects for one, or one for two—that is, they can think about things and suffer from identity confusion. However, it is implausible to attribute them with beliefs about the identity relation.

It is very hard to say how compelling this worry is without a more detailed discussion both of the mental abilities of small children and puppies, and of how cognitively demanding beliefs about identity really are.

The second worry is more evidently compelling. This is that the identity beliefs that the belief theorist claims to be constitutive of confusion are in fact representationally unavailable to a confused thinker, in virtue of the fact that she's confused.

In particular, this is a claim that has been made about thinkers who suffer from *combinatory* confusion (e.g., Millikan 1994, 2000; Camp 2002). The belief theorist says that Elmar's daughter's combinatory confusion consists in the fact that she has the belief *that Bill is identical to Biff*. However, the worry goes, her having this belief would require certain conceptual or representational resources. It would require her to have one, stable concept (or representation) that determinately refers to Bill (and not Biff), and one that determinately refers to Biff (and not Bill). But this is exactly what Elmar's daughter lacks, in virtue of her confusion. She fails to *distinguish* between Bill and Biff and therefore has a single concept (or representation), which she applies indiscriminately. This means she lacks the conceptual or representational resources required to represent Bill (as distinct from Biff), and Biff (as distinct from Bill) such that she could form the false belief (*that Bill is identical to Biff*), which is meant to constitute her confusion.

These worries have led to the conclusion that, in spite of our everyday ways of describing states of confusion, confusion cannot consist in false identity beliefs about the objects of one's thoughts. Rather, what Unnsteinsson calls 'the concept model' starts with the idea that identity confusion consists in a kind of failure to keep track of identity that is more basic than—that is, *antecedent* to—the failure of false belief (Millikan 1994; 2000). In line with this, the central claim of the *concept model of confusion* is that identity confusion consists in corruption of a thinker's object-directed concepts or representations.

What kind of corruption? In the case of combinatory confusion, one has a single concept (or mental representation) whose function in one's mental economy is to represent a single thing. But, that representation is tokened in response to two (or more) distinct individuals, used in the formation of beliefs in response to information acquired from two (or more) distinct individuals, and so forth. In this sense, the concept is informationally corrupted.

The concept model is inspired by Ruth Millikan's (1994; 2000) account of *combinatory* confusion. But a thinker's concepts or representations are not corrupted in the same way when she suffers from *separatory* confusion. However, as Unnsteinsson envisages things, the concept model is an account of both species of confusion. This can sound like a problem for the concept model, but might instead be seen as part of why a concept theorist of traditional (combinatory) confusion will tend not to assimilate combinatory and separatory confusion under a single, mental kind. However, on the extended concept model envisaged by Unnsteinsson, the idea would be that the concepts or representations employed by a thinker who is in a cognitive Frege case are *malfunctional* (though perhaps not informationally corrupted). The thinker has two distinct concepts or representations, though both are tokened in response to the same object, are used to form beliefs in response to information acquired from only one thing, and so forth. In a perfectly functional system, all the information coming in from a single source would be treated as such, and every token thought that represents the same thing would be used as such. A perfectly functional system would be, as Millikan (2000, p. 172) puts it 'nonredundant' (as well as 'nonequivocal').

So, what, according to Unnsteinsson, are the details we can add to the belief model of confusion, which would allow it to avoid the worries that have traditionally motivated the concept model? The central move is that, on Unnsteinsson's belief model, the false identity beliefs that constitute confusion are *strongly implicit*. They are beliefs truly attributed to confused thinkers in order to explain their (inferential and other) behavior, but they are not represented (they are *implicit*) (Unnsteinsson 2022, p. 38). Furthermore, are beliefs that the thinker may not have the *capacity* to represent (they are *strongly* implicit) (Ibid., p. 40 & 55-60).

In turn, the defense of this claim is that it is not arbitrary or unusual to attribute beliefs in order to explain behavior. And, in some cases, a thinker may be unable to explicitly represent that P despite the fact that their behavior can only be explained by attributing them the belief that P (Ibid., p. 40-3). This, Unnsteinsson claims, is the situation in the case of confusion: the false identity beliefs that constitute a thinker being confused are attributed on the basis of their 'total state of (other) beliefs and disposition', in order to explain their behavior (Ibid., p. 42). The fact that Elmar's daughter believes that Bill is identical to Biff is what explains the kinds of inferences she makes, and the way she behaves with respect to Bill and Biff etc., but this belief is not *represented*. By way of further defending this move, one might note that, even if one is thinking in vehicular terms about representation and the propositional attitudes, one will surely need to acknowledge that not everything that counts as a belief needs to be, or *can* be, represented. For Unnsteinsson, therefore, the false identity beliefs that constitute confusion (and, as we'll see shortly, the *true* identity beliefs that constitute its absence) play a role analogous to a behavioral commitment to *modus ponens*: they act as a kind of principle or presupposition in accordance with which a representational system operates or behaves (Ibid., p. 39-40).

Thus, the appeal to *strongly implicit* identity beliefs answers to the traditional worries about the belief model. First, in positing false identity beliefs as constitutive of confusion, Unnsteinsson is not positing cognitively or representationally demanding states involving the representation of identity. So these beliefs can be truly attributed to small children and various non-human animals (Ibid., p. 48). Second, since the thinker lacks the capacity to represent these beliefs, they don't require the kinds of conceptual or representational resources which would rule out attributing them to confused thinkers.

There may be reasons to resist Unnsteinsson's appeal to strongly implicit beliefs, but I will not focus on them here. Instead, in order to show where Unnsteinsson's account goes wrong, I would like to focus on a desideratum on accounts of confusion that most parties agree on, and on the way that Unnsteinsson's belief-model aims to satisfy it. This is that an account of confusion should come with a corresponding account of the kind of capacity, or *success*, with respect to which confusion is a failure: it should come with an account of clear and unconfused thought.

Central proponents of the concept model fulfill this desideratum as follows. Their claim that confusion consists in the corruption or malfunction of concepts is part of a broader view according to which the role of singular concepts (or representations) is to reidentify, or keep track of, individuals. Confusion occurs when there is systematic failure to reidentify or keep track in thought. Unconfused or clear thought consists in success in reidentifying or keeping track in thought—it consists in possessing singular concepts that are each applied only to a single thing, and having only one singular concept for each thing (See Millikan 2000).

Unnsteinsson acknowledges that an account of confusion should come with a corresponding account of clear and unconfused thought, and also defends the belief model by claiming it can indeed offer this. His view is that, just as confusion consists in strongly implicit *false* identity beliefs*, lack* of confusion consists in implicit *true* beliefs about the identity of the objects of one's thoughts. Just as the confused thinker's inferential and other behavior is explained by false, implicit identity beliefs, the unconfused thinker's inferential and other behavior is *explained* by true, implicit identity beliefs.

Take, for example, Lana Lange, who does not suffer from separatory confusion. Lana believes *that Superman can fly*, and also *that Superman works at the Daily Planet*, and she rationally infers *that someone who can fly works at the Daily Planet*. On Unnsteinsson's view, this (inferential) behavior is correctly explained by an implicit and true identity belief on her part *that Superman is identical to Superman*. Essentially, implicit true identity beliefs 'string together' the coreferential beliefs of an unconfused thinker and thereby explain her behavior. Their role is, roughly, that of an unrepresented principle or presupposition, in accordance with which a representational system operates, and does so in good standing. Unnsteinsson signs up for this account in Ch. 2 of *Talking About*.

3. Campbell's Regress Argument

Before we move on, notice that there are, broadly, two obligations that Unnsteinsson's defense of his belief model must discharge. The first is perhaps more immediately salient: Unnsteinsson is on the hook to show that it is indeed possible to truly attribute the kinds of strongly implicit identity

beliefs—false *or* true—that are meant to constitute confusion and its corresponding success. This obligation receives considerable attention in *Talking About*: Ch. 3 is spent arguing that we are not barred from attributing strongly implicit beliefs in the cases in which the belief theorist needs them.

The second obligation is that these strongly implicit identity beliefs must in fact *be able to explain what they are called on to explain*. Unnsteinsson's claim is that we can truly attribute these beliefs, even though they are not represented, on the basis that they explain the behavior (including inferential behavior) of both confused and unconfused thinkers. So the success of the view relies on them in fact being an essential part of an explanation of the relevant behavior.

I will set the first obligation aside here, and worry instead about the second. I will allow that it is legitimate to attribute strongly implicit beliefs like the ones Unnsteinsson's belief theory appeals to, but will argue that this does not suffice to vindicate Unnsteinsson's account. This is because the identity beliefs he posits don't play the role he needs them to in explaining the behavior of confused or unconfused thinkers.

To illustrate this, I turn to an argument, offered by John Campbell (1987). Understood correctly, this argument illustrates why implicit identity beliefs cannot explain the inferential (and other) behavior of either unconfused or confused thinkers. Unnsteinsson's view supplies us with an opportunity to more clearly understand the insight behind Campbell's argument.

First, however, I want to note that Campbell (1987) himself uses the argument to establish the need for Fregean *sense*. As I understand it, however, it does not establish the need for *sense*, but rather a more general conclusion. At heart, the argument illustrates the inadequacy of explanations of inferential behavior that seek to explain it in terms of the agent's possessing attitudes with purely referential contents.[3] By 'referential content', I mean contents that consist of, or encode, the objects, properties, relations, etc. that they are about, and nothing more than that. On my understanding, the argument seeks to establish the need for the attitudes to encode, not merely their referential properties, but also their coreference with one another, and to do so outside of their referential content.

Campbell's argument begins by asking us to consider arguments like Argument 1[4]:

> **Argument 1**
> George Eliot wrote *Middlemarch*
> George Eliot was born in Nuneaton
> The author of *Middlemarch* was born in Nuneaton

This is a formally valid argument. Our question is, which of its features explain its validity—that is, which explain the entitlement to its conclusion on the basis of its premises? Of course, the content of the predications contained in the premises play a role. But, apart from this, the fact that the two tokens of 'George Eliot' in the premises corefer also play an explanatory role. That is, the argument is valid, and it relies on the coreference of its premises to generate its conclusion.

---

[3] That is, in addition to whatever logical skills or knowledge are presupposed by having the ability to rationally infer.

[4] Thanks to Aidan Gray for suggesting this (slightly non-standard) way of presenting Campbell's argument.

However, this cannot be *sufficient* for the argument's validity, or else Argument 2 would also be valid:

Argument 2
George Eliot wrote *Middlemarch*
Mary Ann Evans was born in Nuneaton
The author of *Middlemarch* was born in Nuneaton

This shows that something more than the coreference of the premises explains the validity of Argument 1. But, what? The Fregean will conclude that the validity of Argument 1 is (partially) explained by the sameness of *sense* of the two uses of the name, 'George Eliot'. But we may wish to resist an immediate appeal to sense and try to explain the validity of arguments like these—which rely on the coreference of their premises to generate their conclusions—in a framework that appeals only to referential content.

In line with this attempt, we might suggest that Argument 1 contains a suppressed premise to the effect *that George Eliot is identical to George Eliot*, and that this explains the validity of Argument 1 and the difference between Argument 1 and Argument 2. Some initial confirmation of this suggestion might appear to come from considering Argument 3.

Argument 3
George Eliot wrote *Middlemarch*
Mary Ann Evans was born in Nuneaton
George Eliot is identical to Mary Ann Evans
The author of *Middlemarch* was born in Nuneaton

Argument 3 is valid, and it is also identical to Argument 2, except for the addition of the same identity premise (considered only in terms of its referential content) as the one we suggested to account for the validity of Argument 1. But, if our question is about what explains or accounts for validity, we should now ask, what explains the validity of Argument 3? If our hypothesis is correct, and the validity of Arguments 1 and 3 are explained by the coreference of their premises plus the presence of an identity premise (construed as having only referential content), then Argument 4 should also be valid.

Argument 4
George Eliot wrote *Middlemarch*
Mary Ann Evans was born in Nuneaton
George Eliot is identical to George Eliot
The author of *Middlemarch* was born in Nuneaton

But, of course it is not. This means we must ask: what is the difference between Argument 3 and Argument 4? And, in answer to this question, we can see that the difference is *not* the referential

content of their premises—it is not the presence or absence of an identity premise construed as having only referential content. Indeed, we could iterate the process just discussed and seek to explain the validity of Argument 3 and the difference between it and Argument 4 by adding a suppressed identity premise but, insofar as this premise is considered only in terms of its referential content, it will not make the difference between a valid and invalid argument.

This shows that the validity of an argument with coreferential premises, which relies on this coreference to generate its conclusion, is explained by more than just the referential content of its premises. Rather, valid arguments with coreferential premises, which rely on this coreference to generate their conclusions, rest, on pain of regress, on some encoding of the coreference of their premises that does not take place purely in their referential content.[5]

And, indeed, this is a familiar feature of the languages we operate with. For example, in a formal language, it may be part of the way the system works that repetition of the same individual constant encodes sameness of reference. Similarly, in natural languages, it might be that repetition of the same term type encodes sameness of reference. Campbell's argument shows that this is no *accident*. It illustrates that, in an argument with premises that corefer to an object *a*, it is never *in itself* the presence or absence of a premise whose referential content is *that a is identical to a*, which makes the difference between a valid argument and an invalid argument. For, it is always possible for such a premise to be present—either explicitly, or as a suppressed premise—and for the argument to be formally invalid nonetheless.

Now, in discussing the conclusion of Campbell's argument, people have sometimes spoken of the need to 'take for granted', 'presume' or 'presuppose' coreference fact (Campbell 1987, p. 276; Recanati 2012, 47-50; Goodman 2022; Unnsteinsson 2022, p. 52). And one might, on this basis, be *tempted* to think the moral of the argument is that there is a need for valid arguments to encode, *rather than explicitly represent,* the coreference of their premises. However, a mistake has been made if one takes the important lesson to relying on the contrast between encoding and the *explicit representation* of coreference (I'll emphasise this further in Section 4). The lesson is rather this: Take an argument like Arguments 2, which is invalid despite its referential equivalence with a valid argument like Argument 1. Add as much *referential content* as you want to Argument 2—*including* in the form of premises whose contents state the identity of the referents of the initial premises—and you will not be guaranteed to generate a valid argument. The referential content of the premises of an argument with coreferential premises, which relies on that coreference to generate its conclusion, do not suffice, on pain of regress, to explain the formal validity of the argument. An encoding of coreference *outside of referential content* is required.

Before I discuss the significance of this for Unnsteinsson's view, I'll pause to be explicit about the lesson of the regress argument for the case of mental representation (as we've outlined it so far, it applies to *arguments*). Assume that inferences are the mental analog of arguments, and that the equivalent of formal validity in inferences is manifest rationality from the subject's point of view. The lesson of Campbell's argument for the case of mental representation is therefore this: the fact that coreferential attitudes stand to one another as premises in a manifestly rational inference that

[5] The addition of an identity premise could be counted as an encoding of coreference that takes place within referential content.

exploits their coreference, can never be explained merely a matter of them having any particular referential content (even referential content which is an identity proposition). That is, the inferential (and other) behavior of thinkers with coreferential beliefs cannot be explained by appeal to the content of attitudes—including attitudes with identity propositions as their content—considered only in terms of their referential content.

What could account for the *difference* between a case in which the coreference of two token attitudes *can* be exploited for inferential purposes and one in which it *cannot?* The inferential equivalent of an 'identity premise'—that is, another attitude whose content is an identity proposition—won't suffice. This is because the question of why the thinker is licensed to exploit—or, as Campbell (1987) puts it, 'trade on'—the identity of the referent of *that* attitude (the identity premise) and the previous two (the original premises) would arise. As we've emphasized, the regress is generated by assuming that the content of the attitudes is purely referential, and the heart of the argument is this: add as much of *this* kind of content as you want, and you won't have accounted for the manifest rationality of an inference that relies on the coreference of the attitudes that serve as its premises, and so won't have explained the inferential behavior of a thinker who makes this inference.

4. **A Dilemma for the Belief Theorist**

As we've seen, applied to the mental case, the regress argument establishes the need for the encoding of coreference in the attitudes, outside of their referential content.

Notice that the argument does not establish the need for Fregean sense, because there are different things that could satisfy the requirement of encoding coreference outside of referential content. We *could* posit *sense* as a component of the content of the propositional attitudes (as in Campbell 1987). This would involve positing a second level of content (over and above referential content), which attaches to each attitude individually. But, the requirement could also be satisfied other ways. For example, it could be satisfied by an *irreducible semantic relation* that is part of the 'coordinated content' of clusters or groups of attitudes (as in Fine's 2007 *semantic relationist* view). Or, it could be satisfied by a *formal* (rather than semantic) encoding of coreference in the attitudes. And, this could involve a formal feature that applies to each token representation individually (for example, *sameness of syntactic type*) or an *irreducible formal relation* that attaches to groups or clusters of token representations (as in Heck's 2012 *formal relationist* solution to Frege's puzzle). All of these views can explain explaining the inferential behavior of agents who are not confused (and those who are), because they all involve a commitment to the idea that the attitudes encode coreference properties as well as referential properties.

As it happens, Unnsteinsson himself appeals to Campbell's regress argument in defending his belief model of confusion (Unnsteinsson 2022, p.49-52). From his perspective, his belief theory is not susceptible to Campbell's argument, because the identity beliefs that it posits to explain the behavior of an unconfused thinker are *implicit*: they are posited to explain the relevant inferential behavior but they are not represented. And, to be fair, talk of the need for thinkers to 'assume' or 'presuppose' coreference or identity might suggest that *implicit* representation of identity would fulfill

the need that the argument points to. And this is what Unnsteinsson's view provides. However, as I have already suggested, this overlooks the real point of Campell's argument.

The real point is the need for something additional to referential content encoded in the attitudes, in order to explain the behavior of an unconfused thinker. In particular, the need is for the attitudes (either individually or as groups or clusters) to somehow encode their coreference with one another outside of their referential content. But, this means that the representationally *implicit* nature of the identity beliefs Unnsteinsson posits to explain the inferential behavior of thinkers is beside the point when it comes to avoiding the regress.

To see this, notice that the belief theorist faces a dilemma, which illustrates what's wrong with the claim that *identity beliefs*—either explicit *or* implicit—explain the behavior (inferential and otherwise) of unconfused or confused thinkers.

For *any* belief—explicit *or* implicit—posited as a premise in an inference that exploits the coreference of its premises, we can ask: is there something additional to the referential properties of that belief encoded in it?

If not—that is, if the attitudes have referential content only and don't encode coreference properties in any non-semantic way—then the relevant inferential behavior will not have been explained (and the regress will be up and running). This is also true for an inference that takes an *implicit* attitude whose content is an identity proposition as one of its premises. If that attitude encodes only referential content, then an additional attitude (perhaps another implicit attitude) will be required. And once we've added that additional attitude, the same need will recur again. The addition of identity beliefs with referential content—*even implicit ones*—can't explain the agent being disposed to, and in a position to, rationally make inferences that rely on the coreference of the attitudes that play the role of premises.

If, on the other hand, the implicit belief that's posited does *not* have merely have referential content—that is, if it *does* encode its coreference with other attitudes outside of its referential content—then we are entitled to ask: what motivates the view that we always need an *implicit* belief to explain inferential (and other) behavior in the first place?

Take, for example, Lana Lange, who is *not* in a Frege case. She has the explicit belief *that Superman can fly* and the explicit belief *that Superman works at the Daily Planet*, and she is disposed to (and in a position to) rationally infer *that someone who works at the Daily Planet can fly*. In explaining her inferential behavior, the view we are imagining posits an *additional and implicit* belief *that Superman is Superman*, which does not merely encode its referential properties, but also encodes its coreference with Lana's explicit beliefs. This is a possible position to be sure, but it is unclear why it is preferable to the view that Lana's beliefs *in general* (including her explicit beliefs) encode their coference with one another (through Fregean sense, relational semantic content, sameness of syntactic type, or an irreducible formal relation between token representations). Once we have allowed the *need* for beliefs to encode coreference outside of their referential content, specifically *implicit* identity beliefs that are always needed to string together explicit beliefs, seem otiose. Why not grant the direct encoding of coreference by ordinary beliefs?

Unnsteinsson spends a good deal of time, in *Talking About*, defending the belief model of confused and unconfused thought, by defending the idea that we can attribute strongly implicit

identity beliefs to thinkers in order to explain their inferential behavior. The problem with the belief model, however, is not necessarily that we cannot legitimately posit strongly implicit identity beliefs, merely on the basis that they explain behavior. The problem, by my lights, is that strongly implicit identity beliefs with referential content do not in fact explain the relevant behavior, and implicit identity beliefs with the kind of semantic or formal properties that would mean they *could* explain the relevant behavior are not in fact needed. Once we allow that beliefs can have these sorts of properties, there's no need to posit implicit identity beliefs to explain every case of unconfused thought (or every case of confused thought).[6]

I conclude by noting that, while I have not defended the (standard) concept model of clear and confused thought here, it does come out looking preferable to the belief model, in light of my argument. It is open to concept theorists to claim that coreference is encoded in the propositional attitudes through sameness of concept, construed in either semantic (Fregean) or formal (vehicular) terms. Whether one of these is the *right* way to think about how the attitudes encode their coreference other attitudes depends on questions that go beyond the scope of the current discussion but, insofar as the concept model has resources to explain the behavior of confused and unconfused thinkers, it remains preferable to the belief model, which lacks these resources.

References

Camp, J. L. (2002) *Confusion.* Harvard University Press.

Fine, K (2007) *Semantic Relationism.* Wiley-Blackwell

Goodman, R. (2022) 'Trading on Identity and Singular Thought', *Australasian Journal of Philosophy*, 100: 2, 296-312

Heck, R. K (2012) 'Solving Frege's Puzzle', *Journal of Philosophy*, 109, 132-74. (Originally published under the name 'Richard G. Jeck, Jr.')

Millikan. R. G. (1994) 'On Unclear and Indistinct Ideas', *Philosophical Perspectives*, 8, 75-100.

Millikan. R. G. (2000) *On Clear and Confused Ideas: An Essay About Substance Concepts.* Cambridge University Press.

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

Unnsteinsson, E. (2019) 'Frege's Puzzle is About Identity After All', *Philosophy and Phenomenological Research*, 99: 3, 628-43.

[6] This allows that there are *some* cases where there is reason to posit implicit beliefs (including implicit identity beliefs). This is part of what I have granted for the purposes of my argument.

Unnsteinsson, E. (2022) *Talking About*: *An Intentionalist Theory of Reference*. Oxford University Press.