Frege's puzzle is about identity after all

Elmar Unnsteinsson

Preprint. Please cite published version:
Philosophy and Phenomenological Research (2019) 99(3):628–643
doi:10.1111/phpr.12516

Abstract

Many philosophers have argued or taken for granted that Frege's puzzle has little or nothing to do with identity statements. I show that this is wrong, arguing that the puzzle can only be motivated relative to a thinker's beliefs about the identity or distinctness of the relevant object. The result is important, as it suggests that the puzzle can be solved, not by a semantic theory of names or referring expressions as such, but simply by a theory of identity statements. To show this, I sketch a framework for developing solutions of this sort. I also consider how this result could be implemented by two influential solutions to Frege's puzzle, Perry's referential-reflexivism and Fine's semantic relationism.

1 A puzzle about identity

What exactly does Frege's puzzle have to do with identity or identity statements? Interestingly, according to what Joan Weiner (1997: 269) calls 'the standard interpretation,' there is not supposed to be any connection at all between the two. Michael Dummett held, in his 'What is a Theory of Meaning?', first published in 1975, that the puzzle can be formulated in terms of any 'atomic statement' (Dummett 1993: 24, 86). Richard Mendelsohn (1982: 281-282, 2005: 30) presents a more detailed argument for this claim, which has been repeated or, at least, implicitly

*(✉) elmar.geir@gmail.com

*Thanks to Daniel Harris, Thomas Hodgson, Finnur Dellsén and to members of the audience at the UCC-UCD Workshop in Cork and QUB Mini-Workshop in Belfast, both in 2016. Many thanks also to several anonymous referees for comments and suggestions.

endorsed by many philosophers since (e.g., Fine 2007). In his book on the puzzle, Nathan Salmon (1986: 12) argues that the puzzle 'has virtually nothing to do with identity.' His thought is, roughly, that since the puzzle can easily be stated without any appeal to identity statements it follows that it is really about something else, namely the cognitive significance of referring expressions as such.

The standard interpretation of the puzzle, however, is wrong. I will argue that any statement of the puzzle makes essential reference to statements or beliefs involving the identity relation. Others have argued for this on exegetical grounds—e.g., Weiner (1997), Beaney (1996: ch. 6)—but I am only concerned with the puzzle's form as it rears its head in contemporary debates. Yet, if my argument holds water, we are free to take Frege at his word; the puzzle is about how true, coreferential identity statements can differ so dramatically in their cognitive effects.

It should be noted, before presenting the arguments, that the conclusion will be that posing the puzzle essentially involves statements *or beliefs* about the identity of objects. Some might object that, on a charitable reading, the standard interpretation only says that *statements* of identity are unnecessary. But insisting on the importance of this distinction in the very formulation of the puzzle begs the question. As is well known, it can also be stated in terms of an individual's beliefs or other contentful mental states, and the argument developed below in favor of the standard interpretation can just as easily be given in terms of beliefs. The actual source of the puzzle, even when stated without explicit mention of identity, could very well be some implicit assumption involving beliefs or, simply, potential statements about identity that are made salient in the context as described. And this is exactly what I will argue. Indeed, if the standard interpretation is taken, merely, to hold that no actual statement involving explicit reference to the identity relation need be countenanced, it would be trivial, not warranting its considerable prominence in some of the works cited above. Most importantly, however, such an interpretation, even if true, would not warrant the widespread assumption that solving the puzzle—as opposed to rejecting some intuitions underlying its formulation—calls for a completely general theory of reference and truth-conditional content.[1] A rough analogy would be to say that since the Liar paradox can be stated with 'I am lying now' it is not a paradox about truth and, thus, that it must be about something more general, like predication.

[1]Obviously, there might be other reasons for developing such a theory, the only point here is that Frege's puzzle doesn't provide one in the way most have assumed.

Now, let 'P' be the proposition that Frege's puzzle is not essentially a puzzle about identity. Being essentially about identity is shorthand for: referring essentially to some potential statement or belief involving the identity of objects. The argument *for* P, then, proceeds as follows. Intuitively, there is a semantic difference between (1) and (2).[2]

(1) Tully is an orator if Cicero is.

(2) Cicero is an orator if Cicero is.

It seems, however, that (1) and (2) attribute the same property to the same individual. Both are true if and only if that individual x—the one named both 'Tully' and 'Cicero'—has the property of being an x such that x is an orator if x is. If so, what explains the semantic difference between the two? This is Frege's puzzle and, since there is no mention of identity in (1) or (2), it follows that P is true. The puzzle can, it seems, be stated with any dyadic relation one cares to think of. It can even be stated using only monadic sentences with distinct but coreferring singular terms.[3] For what explains the apparent semantic difference between (3) and (4), given that they have identical truth conditions?

(3) Tully is an orator.

(4) Cicero is an orator.

Thus, P seems unassailable. And it is only natural, on this basis, to hold that the puzzle arises simply by considering the different semantic properties of coreferring singular terms like 'Tully' and 'Cicero'. And, so, extant solutions to the puzzle either deny the intuition that there is such a difference, or they involve the claim that the semantic content of a name is not simply its referent; perhaps it's a Fregean sense of some sort.

Yet, P is demonstrably false. To start, note that many theorists would hold that the puzzle as stated is not well motivated. We need to know, more precisely, why there appeared to be a semantic difference in the first place. Usually, the missing link is provided by claiming that (1) and (2) can convey different information to someone who understands both sentences (e.g., Fine 2007: 34). I assume here that something along these lines is necessary for a full statement

[2]My formulation here is intended to be neutral on the question of the proper bearers of semantic difference. The reader is free to think of (1) and (2) as sentence types, dated utterances, propositions, statements, or something else. Note, however, that this neutrality will be dropped, as it must, when a rough theory is sketched in the next section.

[3]Fine (2007: 52) calls this the monadic version of the puzzle, see §3 below.

of the puzzle.[4] Furthermore, most theorists take the notions of 'conveying different information' or having 'different cognitive effects' to be epistemic, calling for relativisation to an epistemic agent.

So, we need to specify the minimal conditions for being an epistemic agent S such that (1) and (2) have different cognitive effects on S even when S understands both. A natural suggestion is that the condition will have something to do with S's epistemic access to the identity or distinctness of Cicero/Tully. Consider the hypothesis that (1) and (2) have different cognitive effects on S if and only if S is ignorant of the fact that Cicero = Tully. Proving the hypothesis would show that the puzzle essentially involves identity. There are, then, two relevant epistemic states for S.

(A) S believes truly that Cicero = Tully.

(B) S lacks the true belief that Cicero = Tully.

If we assume that belief is closed under logical consequence it follows that, in state (A), (1) and (2) will have exactly the same cognitive effect on S, namely, one will be trivial or uninformative just in case the other is. To illustrate, consider an example of two names or name-like expressions which are generally known by all to be coreferential, like 'John F. Kennedy' and 'JFK'. (5) and (6) will, generally, have exactly the same cognitive effect.

(5) JFK was catholic if John F. Kennedy was.

(6) John F. Kennedy was catholic if John F. Kennedy was.

And the same would apply in the monadic case. This is because, generally, people are aware that John F. Kennedy = JFK.[5] Therefore, Frege's puzzle does not arise for S in epistemic state (A).

[4]Note, however, that Nathan Salmon would disagree. He argues that sentences, relative to context, encode pieces of information eternally. Thus, on his view, the puzzle can be stated directly in terms of differences in information encoded by two sentences. In contrast, I assume here that the information value of a sentence must be relativised, in some way or other, to the speaker's utterance of the sentence on a given occasion. Salmon states his opposing view clearly in 'Two Conceptions of Semantics' (2005). Arguably, however, the notion of 'encoding information' cannot be taken as primitive in stating Frege's puzzle, and this is borne out by how other theorists invariably bring in something like the missing link mentioned here.

[5]Those who are inclined to think that the acronym and the full name are, in fact, one and the same expression, only written differently, should be reminded that 'scuba' is—or was?—also an acronym. The pair in (5) is certainly susceptible to traditional Frege cases.

Now, consider epistemic state (B). This kind of ignorance comes in many stripes but, here, there are three variants that require consideration. First, S's ignorance may be due to the fact that S is unfamiliar with practice of using 'Cicero' or 'Tully' or both. Others have convincingly argued that the puzzle does not arise in a case of this sort, so I will not repeat the argument here (Salmon 1986: 60). Secondly, S's ignorance may consist in cognitive indifference or suspension of judgment about the identity of Cicero/Tully. Here, (1) and (2) will potentially have different cognitive effects on S. Presumably, S will consider (2) trivially true but (1) neither true nor false. In the monadic version of the puzzle, similarly, (3) and (4) would potentially have different cognitive effects on S if S suspends judgment. We cannot, however, predict that (3) and (4) would necessarily occasion, for S, different judgments of truth value. What we can say, still, is that, in virtue of epistemic state (B), it is not necessary that S will accept (3) as true if and only if S accepts (4) as true and, thus, they are potentially different in their cognitive effects. But in state (A) the strict biconditional will hold: necessarily, S accepts (3) iff S accepts (4). This is to say that the antecedent and consequent must have the same truth *condition* and not merely the same truth value.

Finally, S's ignorance may consist in the fact that S lacks the true belief that Cicero = Tully in virtue of *having* the false belief that Cicero ≠ Tully. Of course, the same reasoning as in the second variant applies here too. (1) and (2) will have different cognitive effects on S, in particular, (2) could be considered trivially true while (1) is considered either true or false or neither. The treatment of the monadic version is also similar to before. We are unable to predict that S assigns different truth values to (3) and (4) but, in virtue of being in state (B), it is not necessary that S will accept one as true if and only if S accepts the other as true. This explains their potentially different effects across different possible worlds. Once S is in state (A), however, the strict biconditional is predicted to hold.

It seems, then, that we have identified an epistemic condition which must always be involved in the full statement of Frege's puzzle. Assuming (A) and (B) to be exhaustive of S's relevant state of belief with respect to the identity of Cicero/Tully, we can say that the puzzle arises if and only if S is not in state (A). It should be clear that this argument can be run with any dyadic predicate in place of the conditional in (1) and (2) and, so, it is completely general. It is only prudent to conclude that P is false; Frege's puzzle has everything to with identity. It only arises in cases where the thinker has a false belief about the identity of the object in question. In a possible world where all facts about identity are self-evident and a priori, no one would ever worry about solving this puzzle. One could worry, however, in a world where all conditionals are known a priori.

If this argument is correct it is deeply misleading to say that the puzzle has virtually nothing to do with identity. Yes, on the surface it can be stated in terms of the cognitive difference between sentences or utterances in which identity is not mentioned. But the puzzle is still essentially about identity in that it arises in virtue of properties of the mental state of someone who happens to be ignorant, in the relevant respect, about the identity of an object.

On the standard story, the puzzle is about the semantic or cognitive contribution of singular terms in larger expressions. Names appear to contribute different information even if they are, clearly, coreferential and, so, the question is: What do singular terms contribute, which is different from a mere referent? And this is where Frege, and others, have found a need for dual-component semantics which distinguishes sense or meaning from reference; the former being the semantic or cognitive contribution of a singular term. But any such a solution, of course, remains controversial.

At the most basic level, however, Frege's puzzle is about exactly what Frege said it was about: In virtue of what does 'a = b' seem to differ in cognitive significance from 'a = a' when 'a' and 'b' corefer? Positing senses as the cognitive contributions of any term like 'a' and 'b,' in any context, does indeed count as a possible solution to this problem. But it is important to realise, in light of the argument so far, that the puzzle allows for a much more limited approach, which focuses on the peculiarities of identity as such.

It is not to my purpose to argue for a particular proposal of this kind here. However, in order to rebut a possible objection to the more modest conclusion announced in the title, some rough idea needs to be sketched. Salmon, and others, might well object as follows.

> Suppose that the puzzle, as you understand it, is solved by some theory that applies only to contexts involving identity. Let's, if you like, simply get rid of identity from our language. But then I cannot see why the puzzle would not just resurface in, for example, contexts involving coreferring names related by a conditional, like in (1) and (2). If so, it is clear that the puzzle has nothing essential to do with identity.

But the objection is subject to a similar argument as the one given against P before. We must assume that the agent S for whom (1) and (2) differ in cognitive effect either knows that 'Tully' and 'Cicero' corefer in (1) or S doesn't know. The difference only arises if S lacks the true belief in question. Now, of course, it is

strictly possible to articulate S's ignorance without using the symbol for identity: S does not believe that 'Cicero' and 'Tully' corefer. But the notion of coreference is naturally understood as the identity of referents and, so, the objection loses its bite. The point here is not that the sentences "'Cicero' and 'Tully' corefer" and "'Cicero' = 'Tully"' have the same meaning or content, so the latter can be analysed in terms of the former. It is, rather, that any theory of the coreference of names will, at a minimum, assume that it partly consists in the fact that the referents of the names stand in the identity relation to one another. Thus, the puzzle only arises if identity or something that presupposes it is reintroduced into the language.

Another objection that merits explicit mention is as follows. Referring to Jennifer Saul's (1997, 2007) work, someone might insist that identity confusion is not required to get differences in cognitive effects or intuitive truth conditions for sentences like (5) and (6), far from it. For example, we know full well that Clark Kent is Superman but still (7) and (8) seem importantly different.

(7) Clark Kent went into the phone booth and Superman came out.

(8) Superman went into the phone booth and Superman came out

The same reasoning, it seems, applies to this pair:

(9) I visited St Petersburg once, but I never made it to Leningrad.

(10) I visited St Petersburg once, but I never made it to St Petersburg.

As Saul argues, the intuitive truth conditions of (7) and (9) appear to change significantly after the substitution of allegedly coreferring names in (8) and (10). The cognitive significance of the two pairs seems correspondingly different. Most importantly, nothing seems to hang on the speaker or hearer being confused about the identity of the relevant objects.

This line of argument is part of the reason why I used the pair of 'JFK' and 'John F. Kennedy' instead of the more traditional examples involving Superman or Cicero. This is a better, more realistic example, for 'JFK' and 'John F. Kennedy' are both part of the active public language vocabulary of many speakers. If one is competent with either it is fairly likely that one is competent with the other. 'Tully' is so unfamiliar to most that the relevant identity statement can too easily sound like the introduction of a new name. 'Superman' and 'Clark Kent' evoke role-based interpretations too easily. So, (7) can be paraphrased as 'Clark Kent went into the phone booth and came out as/in the role of Superman.' 'St Petersburg' and 'Leningrad' readily evoke historical period-based interpretations.

Thus, (9) could be, roughly, 'I visited St Petersburg once, but only when it was so-called' (also denying having been there when it was named Petrograd).

Certainly, this difference between the JFK-case and the others will not matter when the dialectic is already assumed to be about the semantics of referring expressions as such, in any context. And my aim is not to show that such a project has no justification, only that Frege's puzzle doesn't give us one. The current argument is aimed at isolating precisely the factors giving rise to the puzzle in the first place. The argument based on Saul's examples suggests that the crucial factor is (i) the different semantic contributions of coreferring proper names, while I have been trying to show that it is (ii) the fact that the speaker or hearer is confused about the identity of the object referred to. So, it stands to reason that my objector, to avoid begging the question, needs to provide a case where the truth conditions are intuitively different, as in the pairs above, while the names at issue do not readily evoke role-based or period-based interpretations (and these might be the referents of some names for all I have said so far). However, if the argument against P is correct, this would seem to be impossible.

(11) John F. Kennedy went into the phone booth and JFK came out.

I take it that (11) will be judged a little odd, but that it would probably be taken as an awkward way of saying that John F. Kennedy went into the phone booth and then came back out. Otherwise, uttering (11) would suggest that the speaker is not aware that John F. Kennedy is JFK. It's hard to see what it would mean for John F. Kennedy to come back out *in the role of* JFK.

2 Framework for limited solutions

Even if the argument so far is sound, it is not immediately clear why it matters. Surely, the identity puzzle might still be solved by a theory of the cognitive differences between coreferential names. And if it turns out that there is no plausible solution which limits its scope to identity statements—i.e. doesn't generalise to the semantics of all other occurrences of names—the argument would seem of little consequence. In this section I argue, however, that such a limited answer to Frege's puzzle is at least possible, without resorting to something as far-fetched as ridding our language of the symbol for identity.

To start, let's briefly introduce two notions. First, there is the idea of presupposition. A proposition is presupposed when a speaker takes it for granted in making an utterance. For example, in uttering,

(12) I need to pick up my sister,

I would merely presuppose, and not assert, that I have a sister. Call this proposition p. Sometimes I would presuppose that the hearer also knows p, but I can also presuppose that the hearer will accommodate, i.e. come to accept p by recognising that p is presupposed. Presuppositions typically survive embeddings—they are 'projective'—but then it appears that they sometimes become cancellable. So, if (12) is embedded in 'It's not the case that …,' p survives but would be canceled by adding, for example, 'In fact, I don't have a sister.' Such a cancellation, when added to (12), would be infelicitous.

Secondly, we will say that a proper name is 'in play' in a given conversation if (i) it is being uttered or (ii) it is contextually salient. For simplicity, in what follows, I ignore utterances where a name is merely mentioned. Contextual salience is achieved in a variety of different ways. The name may have been uttered before in the conversation or have obvious associative links to other things that were said or to something in the perceptual environment of those conversing. So, for example, if two people are talking generally about former presidents of the US, this may significantly raise the contextual salience of their names—at least the most familiar ones—even before any name has been uttered. The notion is similar to what Craige Roberts (2004) calls 'weak familiarity' in her account of the semantics of pronouns.

Now I will give two arguments for the following thesis, call it 'R':

> If names $\ulcorner n_1 \urcorner$ and $\ulcorner n_2 \urcorner$ are in play in a conversation between S and H, and S utters a sentence containing either name, S will normally presuppose that $\ulcorner n_1 = n_2 \urcorner$ or that $\ulcorner n_1 \neq n_2 \urcorner$.

Note that the notion of presupposition assumed in R is supposed to be neutral as to the triggering mechanisms. That is to say, either the relevant presupposition is semantically triggered by the use of a name or presupposition is only a pragmatic phenomenon, and not a proper part of the semantics of names. The argument to be developed works on either assumption. First, consider sentences with two names.

(13) JFK admired John F. Kennedy.

In uttering (13), the speaker would normally presuppose—falsely—that JFK $\neq$ John F. Kennedy. Both names are brought to play by being uttered. And the presupposition seems to pass standard diagnostic tests. 'JFK didn't admire John

F. Kennedy' will carry the same presupposition and make it cancelable by, for example, adding 'In fact JFK was Kennedy and he loathed himself.'

Secondly, consider a whole conversation between S and H in which 'JFK' and 'Kennedy' are being used by both to talk about the 35th president, although neither utters a sentence, like (13), with both names. Now, say S utters (14).

(14) JFK was presidential.

It is generally assumed that (14) would presuppose that JFK existed (van der Sandt 1992) and, further, it is often argued that the presuppositional profile of a name, like 'JFK,' mirrors that of definite NPs (Geurts 1997; Hawthorne & Manley 2012; Fara 2015). Like definite descriptions, then, (14) would normally presuppose that there is exactly one JFK, of which the speaker then says that he was presidential. This uniqueness presupposition can be formulated as having the content: JFK is distinct from everyone else and identical to himself. In particular, the speaker will take for granted, in uttering (14), that JFK is distinct from other individuals whose names are in play in the conversation. It may take some inconsistency or error to raise this presupposition to salience. Suppose, for example, that later in the conversation, S utters (15), but without any prosodic or gestural indication that this amounts to a change of mind since (14) was uttered.

(15) John F. Kennedy was not presidential.

Realising the error, H may respond: 'You're assuming JFK and Kennedy are different people. But JFK *is* Kennedy.' For instance, further embedding (15) in the antecedent of a conditional would carry the same presupposition but, as before, it would be cancelable. In this context it seems to be canceled, for example, by 'If (15), then I'm a Dutchman: JFK and John F. Kennedy are one and the same person!'

Both of these arguments use examples where the speaker presupposes (falsely) that $\ulcorner n_1 \neq n_2 \urcorner$, but corresponding cases are easily constructed for the presupposition that $\ulcorner n_1 = n_2 \urcorner$. Thus I conclude that thesis R has some initial plausibility. R articulates the idea that competent users of a name will, in uttering the name on a given occasion, represent themselves as not being confused about the identity of the object to which they thereby intend to refer. Speakers presuppose that they have no relevant false beliefs about the identity or distinctness of the object.

Of course, something must be said about cases where speakers explicitly speculate or wonder about identity. This, I believe, can be done (see Unnsteinsson 2016). My purpose here is not, however, to make a rock-solid case for R,

but to articulate a limited response to Frege's puzzle that is worthy of consideration. Based on what has been argued so far, we can say that if speaker S is not in epistemic state (A) with respect to two names, then S's utterance of either name may, given certain conditions, involve presupposition failure. And Frege's puzzle arises only for speakers who are not in a state of kind (A). It follows that speakers who are susceptible to Frege cases will be liable, in virtue of their ignorance of identity, to presupposition failure. Surely, there is a variety of opinion about what happens when presupposition fails and about the extent to which it spells trouble in semantics and pragmatics, e.g., whether it gives rise truth-value gaps or not.

But even on pragmatic theories of presupposition, where failures need not lead to semantic catastrophe, false presuppositions about identity may be problematic and, perhaps, impugn a speaker's communicative competence as a user of a name (cf. Yablo 2006: 167–168). On a Stalnakerian theory, the confused speaker would, in uttering the name, presuppose something that is not actually in the common ground (Stalnaker 1999). In our examples above, the speaker presupposes that JFK $\neq$ John F. Kennedy, which is not in the common ground. In such a situation, then, a speaker's use of a proper name is, to this extent, infelicitous and subject to valid correction or criticism. So, to be precise, the limited solution to Frege's puzzle clings to the rather counterintuitive claim that failing to make the 'correct presupposition' in accordance with R makes speakers pragmatically incompetent users of the name(s) in question. Note, however, that the counterintuitiveness is mitigated by requiring that both names, say 'Superman' and 'Clark Kent' are in play in the relevant situation. So, if Lois Lane is in no way inclined, at a given time, to talk about Superman/Clark Kent with one of the names but only the other, she can count as a pragmatically competent user of that name at that time. Otherwise, she might be inclined to say such things as 'Superman is bigger than Clark Kent,' which, on the theory under consideration, would make her pragmatically incompetent, even if she is not aware of this herself.

A limited solution of this kind needs, then, support from a theory of pragmatic competence, where this is distinguished from syntactic and semantic competence. The theory would describe the mental mechanisms in virtue of which speakers and hearers can form communicative intentions and infer a speaker's meaning on the basis of semantics and other features. On some accounts this requires a dedicated mindreading module. The details will be left to one side here (Carston 1998; Neale 2005; Sperber & Wilson 2002). Note, however, that there is good reason, on this approach, to abstract away from the type of ex-

ample made classic by Frege himself. I have in mind examples where a whole linguistic community makes a significant discovery about the identity of an object, using names already familiar to everyone, like 'Hesperus' and 'Phosphorus'. When absolutely everyone is 'incompetent' with an expression, and discovering the reason for the error is in principle difficult—e.g., because telescopes haven't been invented yet—there is a sense in which no one could be incompetent with the expression. It is the sense according to which perfect uniformity of use in a group is sufficient for competence. A limited solution to Frege's puzzle would, at least, require a stronger notion of pragmatic competence.

If this rough sketch of a theory has some plausibility, it shows that Frege's puzzle about identity might be approached directly by a theory of statements and beliefs about the identity or distinctness of objects. What I have presented is merely a framework for such an approach, mentioning some of the junctures at which more development is called for. But the upshot is, most immediately, that theorists have been wrong to proceed as if a complete theory of the semantics of names, or referring expressions more generally, must be in place before Frege's puzzle can be dispensed with. Rather, theorists are within their rights to refrain from generalising a solution to the puzzle to uses of names or referring expressions in contexts not involving identity. They would claim that speakers who are confused about the identity of an object tend to make false presuppositions which defeats their status as fully competent users of the corresponding linguistic expression. Consequently, exactly those uses of identity statements that have interested philosophers for so long would be classified as, strictly speaking, infelicitous. The person for whom the sentence would be cognitively significant is not a pragmatically competent user of the names in question.

3 Relationism and reflexivism

Now, consider how two fairly detailed, recent solutions to Frege's puzzle could be altered so as to address only the problem of identity statements outlined above. Significantly, the alterations are easily applied and the resulting views not without their merits. First, John Perry (1988, 2012) has developed the reflexive-referential theory which is intended to capture all types of referring expression: names, pronouns, indexicals, demonstratives, definite descriptions, and so on. For simplicity, we focus only on proper names here. Very roughly, the idea is that speakers, by uttering a sentence containing a name, thereby convey at least two types of proposition. One is the singular proposition containing the worldly

object referred to, the truth condition of which depends on the actual state of that very object, even when evaluated at a different possible world. On Perry's view, however, speakers will also convey various kinds of reflexive propositions. These are also singular proposition, but they always contain the act or event of uttering some specific expression whereby the referential proposition was expressed. Thus, when I utter

(16) JFK was in command of the PT-109 in WW2,

the referential proposition I express is simply that JFK was in command of the PT-109 in WW2, the truth condition of which depends on facts about Lieutenant Kennedy. I also convey a reflexive proposition about my very utterance of the expression 'JFK' in the act recorded by (16), namely the proposition that the person referred to by 'JFK' in that utterance was so and so. To specify this proposition completely however, we need to say it is the proposition

(17) that the person the convention exploited by (16) permits one to designate with 'JFK' was in command of the PT-109 in WW2.

According to reflexivism, then, utterances of sentences with distinct but coreferring names will differ in cognitive significance, not in virtue of differences in the referential propositions expressed, but in virtue of differences in reflexive propositions like (17) (Perry 2012: 122).

There are plenty of good reasons to posit reflexive propositions, as it seems clear that speakers intentionally convey all sorts of information about their own utterances when they communicate. As Perry shows, conveying such information can form an important part of a speaker's overall plan in getting something across. But solving Frege's puzzle is not a good reason to posit reflexive propositions.[6]

To see this, consider a variation on a case presented by Joseph Camp (2002: ch. 3). Frida decides to order an ant farm from the hobby store and emails the shopkeeper to explain what kind she would like. She wants to have a bunch of small ants and two bigger ants. The shopkeeper is happy to oblige and sends her the package by mail. Even before the ant farm arrives, Frida has decided that she's going to call both of the big ants by the name 'Joe,' that way she won't have to remember two names, she thinks. By mistake, she receives an ant farm with lots of small ants but only one big ant. Frida never realizes, however, and starts

[6]Buchanan's (2014) theory is similar to Perry's in many respects and, thus, faces a problem akin to the one mentioned here (cf. Peet 2016; Unnsteinsson 2018b).

calling 'both' of her ants Joe, finding it only slightly odd that she can only ever see one of them at any given time. But she believes they don't like each other. She also imagines various distinguishing marks which she thinks she can use to separate the two Joes. She might even say things like, 'Hi there Joe! Joe was here just a few minutes ago. Of course I mean the other Joe!'

Frida's father finally finds out about the mistake when cleaning the ant farm one day. His daughter has made many drawings of the two Joes, always thinking it absolutely obvious how they can be distinguished by appearance if not by name. He tries to tell her: 'There is only one big ant in your ant farm.' But, naturally, she understands this to mean that one of the two is missing. Frida understands when her dad says, perhaps gesturing at two different drawings of Joe,

(18) Joe is Joe.

It's clear, then, that this particular utterance of (18) is informative while, in a different context, (18) could have been uttered by Frida herself without being anything more than an expression of the trivial truth that a particular ant is self-identical. So, we have yet another formulation of Frege's puzzle. Can reflexive propositions save us? No, they can't. The relevant reflexive proposition conveyed by this utterance of (18) is no less trivial, since the convention exploited by the first occurrence of 'Joe' in (18) is the same convention as the one exploited by its second occurrence. And this is something Frida has always known. On Perry's view, naming establishes 'permissive' conventions (2012: 117) And, in this case, the convention is one according to which 'Joe' *may* be used to designate either one of two ants. The two ants just happened to be one. Thus, it seems, (18) can be informative even when both the reflexive and the referential propositions expressed are undoubtedly trivial.[7]

At this point, Perry might appeal to a different kind of proposition, namely what he calls a 'connected reflexive content' (2012: 108–111). He proposes a distinction between the reflexive contents of utterances or statements and the

[7]Perry's account of Kripke's Paderewski puzzle, which reflexivism seems to handle adequately, doesn't seem to help here since there is only one permissive convention at issue and this is known by the hearer, Frida. It is possible that Perry would argue that 'Joe' doesn't refer to anything at all in Frida's idiolect because the history of the use of the term ends in what, following Donnellan, he calls a 'block'. This means that Frida's uses of the term, from the beginning, failed to identify a referent (Perry 2012: 169). Intuitively, of course, it seems like there is an ant such that Frida is at least sometimes successful in referring to it by 'Joe'. See also the discussion of 'network content' below.

reflexive contents of *beliefs*. At least in forming perception-based beliefs, Perry argues, we construct temporary notions to keep track of different parts of the underlying perception. He calls these notions 'buffers.' To use Perry's example, when I assert that *this dog* is *that dog*, pointing to the same dog twice—partly occluded by a pillar, say—the belief I express is grounded in two different perceptual buffers of the same dog. Now, the reflexive content of the belief is a content that refers to those very buffers, saying, roughly, that each is about the same object. Finally, Perry suggests that when a statement or utterance expresses a belief, its connected reflexive content will be identical to the reflexive content of the belief itself. This new category of content is well suited, he argues, to help to explain complicated Frege puzzles involving demonstratives, especially the 'two tubes'-puzzle proposed by David Austin (1990).[8] Unfortunately, it is not obvious how to apply this kind of solution—explicitly designed for the case of demonstratives—to a case like (18) above. If (18) need not be the expression of a perception-based belief it is unclear what constitutes the beliefs' reflexive content. More generally, the reflexive content of an utterance is relatively intuitive, but the corresponding notion for belief is more difficult to grasp, unless some substantive metaphysics of belief is taken for granted.

But Perry's theory is both rich and complicated and my point here is not to establish conclusively that it fails. In particular, it seems like he could also appeal to what he calls 'network content,' which he takes to 'support' permissive conventions for names (2012: 179). The individual uses of a proper name form a network, growing like a tree from the original use, possibly forming disjointed branches in all directions. If the network happens not to originate in an actual object, the associated name and convention are empty, and this empty 'network content' gets promoted to the status of official or referential content of the relevant utterance (2012: 189–190). Perhaps this is what happens in the ant farm-puzzle above. If so, the reflexivist appears to agree with the idea that Fregean puzzles give rise to something like pragmatically incompetent or corrupt uses of names.

It seems clear, anyway, that this statement of the puzzle depends on the fact that Frida, before the correction, believes (falsely) that Joe is not Joe, as she might have put it. And even if reflexivism fails to solve the puzzle as intended, it has ample resources to provide a limited solution similar to the one above. And this is the main point of my argument here. Eschewing talk of presupposition, the

[8] Fine (2007: 36) develops a very similar puzzle, as have many others (see Pryor 2016: 331–332, also note 35 and references therein).

reflexivist would simply posit a set of reflexive propositions conveyed which, taken together, amount to the speaker conveying the information that there is a unique object referred to by the name uttered, thus an object distinct from any other object salient in the context. There is no need to posit reflexive contents for beliefs or network contents. For example, usually when Frida talks about Joe she intentionally conveys the reflexive proposition that the ant she designates with 'Joe' on a given occasion is not identical to that other ant she also calls 'Joe'. We can suppose that she assumes the context always makes clear which Joe is in question. Having close friends or relatives with the same name easily creates these sorts of situations, for example. At this point one could argue that conveying a false reflexive proposition of this sort, namely a false proposition about identity, impugns the speakers pragmatic competence with the name. The details are not important here, since there are clearly many different ways to cash out the general idea. Of course, there will still be a sense in which reflexivism constitutes a theory of the semantics of referring expressions as such but, in this reconstruction, the theory is not motivated by Frege's puzzle in the way in which this is customary in philosophy of language.

Secondly, Kit Fine (2007) has developed the theory of semantic relationism. Roughly, he proposes to explain differences in cognitive significance between utterances like (19) and (20) in terms of 'semantic requirements' on coreference.

(19) Cicero is Cicero,

(20) Cicero is Tully,

A semantic requirement is, simply, implicit or tacit knowledge speakers are required to possess in order to be semantically competent with a given linguistic expression (Fine 2007: 49–50). On this view, it is a semantic requirement that 'Cicero' refers to Cicero/Tully and it is a semantic requirement that 'Tully' refers to Cicero/Tully, but it is not a semantic requirement that the two occurrences of 'Cicero' and 'Tully' in (20) corefer. It is, however, a semantic requirement that the two occurrences of 'Cicero' in (19) corefer and this means, Fine explains, that these occurrences are strictly coreferential or 'coordinated.' In (20) there is only accidental coreference, however, and no coordination. So, to solve Frege's puzzle, Fine suggests, we can point to a genuinely semantic difference between (19) and (20), namely that only the former expresses a coordinated proposition, while insisting that there is no intrinsic semantic difference between the two names 'Cicero' and 'Tully.' The names differ only in their extrinsic relations of coordination.

The merits of this solution are hotly debated and I won't enter the fray on that question here (Heck 2014; Lawlor 2010; Salmon 2012; Soames 2010; Pickel & Rabern 2017). From the present perspective, relationism is most interesting because it calls for different solutions to dyadic and monadic versions of Frege's puzzle. The dyadic puzzle is posed by using sentences, like the two above, with two name-occurrences. In such cases, it's plausible to think that the relevant expressions are either coordinated or not coordinated by the semantics of the language. But, as Fine himself asks, what about the monadic version? The Fregean will say that there is a difference in cognitive significance between (3) 'Tully is an orator' and (4) 'Cicero is an orator,' but there aren't any other name-occurrences in these sentences to which coordination can be established or not. The same problem arises when such sentences occur in the context of attitude ascriptions. Fine's own response is to hold, still, that there is a *relative* difference between (4) and (3), because each bears different semantic relationships to 'other sentences' (2007: 52). Without having an account of where these other sentences come from, this proposal is difficult to evaluate (see Pinillos 2015 for some discussion).

Thinking of relationism as a limited solution to Frege's puzzle brings some clarity to these issues. First, it has already been shown, in effect, that there cannot be any essential difference between the monadic and dyadic versions of the puzzle. Briefly, if we are to make sure that the notion of cognitive significance doesn't simply reduce to the idea that different names—or the same name uttered on different occasions—can elicit different ideas and mental associations, there is no way to pose the puzzle unless we assume that the thinker in question is confused about the identity of the relevant object. And this applies to the monadic version of the puzzle as well. No one denies that 'Cicero' and 'Tully' could evoke different associations in (4) and (3) and there is no mystery about why this is so. The cognitive difference we are interested in depends, I have argued, on the assumption that the hearer responds differently to the two sentences in virtue of not knowing that Cicero is Tully.

I also argued that by defining a notion of names being in play in a conversation, identity confusion of this sort can be brought to salience, such that a speaker's pragmatic competence is in jeopardy. Of course, the lack of competence doesn't always matter for the practical purposes of communication and, so, it won't be noticed until someone thinks it is relevant. Now, consider the fact that Fine, in solving the dyadic puzzle, has already committed himself to a notion of *semantic* competence that requires recognition of strict coreference. Presumably, he thinks of the speaker who doesn't know that the two occurrences of 'Cicero' in (19) are strictly coreferential as, to that extent, semantically incompe-

tent with respect to the whole sentence (19). But how does relationism extend this prediction to monadic sentences like (4) and (3)? Better yet: How could the semantic relationship between (4) and (3) explain their different cognitive significance when, say, (4) is used but (3) has no contextual salience whatsoever and plays no relevant role in the speaker or hearer's mental life?

The proposal here is that we must confine attention to contexts where two names for the same object are in play. Then it can be argued that, for this more narrow set of Frege cases, the speaker can lack complete pragmatic competence with a particular name. Otherwise, responding to the questions above on behalf of the relationist would have to involve coordinating links between every single use a speaker has made of the name, and even potential or possible ones. It even seems hard for the relationist to stop there, for it seems to be a semantic requirement, also, that speakers recognize coordinating links between their uses of a name and other peoples' uses. This seems psychologically unrealistic.[9]

To conclude, reflexivism and relationism both share important features with the minimal framework for solutions proposed here. Reflexivism posits reflexive propositions in addition to referential ones and their theoretical role is quite similar to what I have proposed as presuppositions to the effect that one knows which contextually salient names name identical objects and which name distinct objects. Reflexivism has the advantage of not bringing in the notion of presupposition, which many would object to, but this simply shows that that notion is not an essential part of the limited solution. It may seem like reflexivism also has the advantage of not making the solution depend on a particular theory of the competence/performance distinction, but this might be doubted, as we have seen. Making such a distinction does appear to be an essential feature of the limited framework. And this is also essential to Fine's version of relationism. But I have shown, in addition, that taking on the assumption that to solve Frege's puzzle one only needs a theory of the nature of identity statements or identity beliefs makes it easier to control what we might call the relationist reverberation effect. To address the monadic puzzle, semantic relations must reverberate through the entire web of name-based propositional attitudes in a linguistic community. Instead, we can keep Fine's idea that dyadic and polyadic sentences pose very specific semantic (or pragmatic) requirements on a speaker's competence. Such sentences can then come into play, in the manner already suggested, even when any actual utterance only involves a monadic sentence. Spelling this pos-

[9] On this point, it seems that so-called formal relationism may be better suited to implement a limited solution to Frege's puzzle than Fine's semantic relationism (Heck 2012, 2014).

sibility out in full detail within a relationist semantics is a project for another occasion, however.[10]

4 Conclusion

I have argued that Frege's puzzle can, in principle, be solved by a theory which concerns itself only with names, or other referring expression, as they occur in statements of identity or distinctness or in beliefs about the identity or distinctness of objects. The moral is not that the project of giving a semantics for referring expressions is ill-grounded, only that, contrary to very entrenched ideas, Frege's puzzle does not provide such a ground. To solve the puzzle is, most directly, to explain the peculiar features of mental states or utterances involving relations of identity or distinctness of objects. I have argued that such a solution need not wait for the development of a complete theory of reference.

Bibliography

Austin, D., 1990. *What's the Meaning of 'This'?: A Puzzle About Demonstrative Belief*. Cornell University Press.

Beaney, M., 1996. *Frege: Making sense*. Duckworth.

Buchanan, R., 2014. "Reference, Understanding, and Communication." *Australasian Journal of Philosophy*, 92(1):55–70.

Camp, J. L., 2002. *Confusion: A study in the theory of knowledge*. HUP.

Carston, R., 1998. "The semantics/pragmatics distinction: a view from relevance theory." *UCL Working Papers in Linguistics*, 10:1–30.

Dummett, M. A. E., 1993. *The Seas of Language*. Oxford University Press.

Fara, D. G., 2015. "Names are predicates." *The Philosophical Review*, 124(1):59–117.

Fine, K., 2007. *Semantic relationism*. Wiley-Blackwell.

Geurts, B., 1997. "Good news about the description theory of names." *Journal of semantics*, 14(4):319–348.

Hawthorne, J. & Manley, D., 2012. *The reference book*. OUP.

Heck, R. G., 2012. "Solving Frege's Puzzle." *Journal of Philosophy*, 109(1-2):132–174.

———, 2014. "In Defense of Formal Relationism." *Thought: A Journal of Philosophy*, 3(3):243–250.

Lawlor, K., 2010. "Varieties of Coreference." *Philosophy and Phenomenological Research*, 81(2):485–495.

Mendelsohn, R. L., 1982. "Frege's Bergiffsschrift Theory of Identity." *Journal of the History of Philosophy*, 20(3):279–299.

———, 2005. *The Philosophy of Gottlob Frege*. Cambridge University Press.

[10] My own implementation of a minimal solution, which differs in important respects from both relationism and reflexivism, is to be found in Unnsteinsson (2016, 2017, 2018a).

Neale, S., 2005. "Pragmatism and binding." Z. G. Szabó (ed.), *Semantics versus pragmatics*, Clarendon, pp. 165–285.

Peet, A., 2016. "Referential Intentions and Communicative Luck." *Australasian Journal of Philosophy*, 95(2):379–384.

Perry, J., 1988. "Cognitive Significance and New Theories of Reference." *Noûs*, 22(1):1–18.

———, 2012. *Reference and reflexivity*. CSLI Publications. 2nd ed.

Pickel, B. & Rabern, B., 2017. "Does Semantic Relationism Solve Frege's Puzzle?" *Journal of Philosophical Logic*, 46(1):97–118.

Pinillos, N. Á., 2015. "Millianism, Relationism and Attitude Ascriptions." A. Bianchi (ed.), *On Reference*, OUP, pp. 322–334.

Pryor, J., 2016. "Mental Graphs." *Review of Philosophy and Psychology*, 7(2):309–341.

Roberts, C., 2004. "Pronouns as Definites." M. Reimer & A. Bezuidenhout (eds.), *Descriptions and Beyond*, Clarendon Press.

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

———, 2005. "Two Conceptions of Semantics." Z. G. Szabó (ed.), *Semantics Versus Pragmatics*, Oxford University Press, pp. 317–328.

———, 2012. "Recurrence." *Philosophical Studies*, 159(3):407–441.

Saul, J. M., 1997. "Substitution and simple sentences." *Analysis*, 57(2):102–108.

———, 2007. *Simple sentences, substitution, and intuitions*. OUP.

Soames, S., 2010. "Coordination Problems." *Philosophy and Phenomenological Research*, 81(2):464–474.

Stalnaker, R., 1999. *Context and content: Essays on intentionality in speech and thought*. OUP.

Unnsteinsson, E., 2016. "Confusion is Corruptive Belief in False Identity." *Canadian Journal of Philosophy*, 46(2):204–227.

———, 2017. "A Gricean Theory of Malaprops." *Mind and Language*, 32(4):446–462.

———, 2018a. "The Edenic Theory of Reference." *Inquiry : An Interdisciplinary Journal of Philosophy*, Online First:1–33. Doi: 10.1080/0020174X.2018.1446050.

———, 2018b. "Referential Intentions: A Response to Buchanan and Peet." *Australasian Journal of Philosophy*, Online First:1–6. Doi: 10.1080/00048402.2018.1432666.

Van der Sandt, R. A., 1992. "Presupposition projection as anaphora resolution." *Journal of semantics*, 9(4):333–377.

Weiner, J., 1997. "Frege and the Linguistic Turn." *Philosophical Topics*, 25(2):265–288.

Wilson, D. & Sperber, D., 2002. "Pragmatics, modularity, and mindreading." *Mind & Language*, 17(1–2):3–23. Repr. in Wilson & Sperber (2012), pp. 261–278.

———, 2012. *Meaning and relevance*. CUP.

Yablo, S., 2006. "Non-Catastrophic Presupposition Failure." J. J. Thomson & A. Byrne (eds.), *Content and Modality: Themes From the Philosophy of Robert Stalnaker*, Oxford University Press, pp. 164–190.