



# Belief Puzzles as Paradoxes of Identity

Ramón García Moya (✉)

BIAP/LOGOS, Department de Filosofia, Universitat de Barcelona, Barcelona, Spain  
ramongarciamoya@gmail.com

**Abstract.** This paper approaches traditional puzzles about belief and belief attributions as if they involved instances of paradoxes of identity. I shall argue that the solution to these puzzles comes with a proper understanding of the way we identify individuals in situations where their *persistence conditions* allow for a “split” (through time or possible worlds) and the way context constraints how we talk about them. My aim in this work is to outline the basics of such a solution and show how well-motivated it is compared to more conventional alternatives.

**Keywords:** Frege’s puzzle · Belief reports · Paradoxes of identity · Derive context

## 1 Introduction

The central puzzle I shall be concerned with in this paper is *Frege’s puzzle*, also known as *the problem of substitution between co-referential names in belief contexts*. The problem is, in a nutshell, to explain how pairs of reports like (1) and (2) differ in truth-value even though standard assumptions about the semantics for belief and proper names disallow any difference in the propositions they semantically express:

- (1) The Greeks believed that Hesperus shines in the evening
- (2) The Greeks believed that Phosphorus shines in the evening

Apparently, ‘Hesperus’ and ‘Phosphorus’ were the names that the ancient Greeks gave to (the personifications of) the brightest heavenly body visible in the evening and the brightest heavenly body visible in the morning, respectively. They initially thought that they were two distinct objects. It was only after they accepted the Babylonian theory that they came to believe that these objects were one and the same and gave it the name ‘Aphrodite’. Standard intuitions say that an utterance of (1) is true whereas an utterance of (2) is false, or so when produced to answer the question <<What did the ancient Greeks believe before accepting the Babylonian theory?>>.

Now, a subject’s beliefs are true (or false) and thus have representational content. On the simplest hypothesis, the content of a belief is individuated by the set of circumstances supporting its truth. On the assumption that truth-supporting circumstances are possible worlds, belief contents become *propositions* as (the characteristic function of) possible world sets. The content of a subject’s belief system will therefore be given by the class

of possible worlds supporting the truth of each of the subject's beliefs. This class is the subject's "doxastic alternatives", and is identified with the set of worlds  $W$  such that the subject believes nothing that rules out the hypothesis that  $w$  is the actual world, for all  $w \in W$  [1, p. 27].

Given the identification of belief contents with propositions, it becomes natural to think of (1) and (2) as characterizations of the Greek's belief system by means of the propositions expressed by their respective prejacent. Thus, if  $W$  is the set of worlds modeling the content of the Greek's actual beliefs, an utterance of (1) is true if, and only if, each world in  $W$  supports the truth of the proposition which 'Hesperus shines in the evening' expresses. Likewise, an utterance of (2) is false if, and only if, at least some world in  $W$  does not support the truth of the proposition which 'Phosphorus shines in the evening' expresses.

However, the occurrences of the names 'Hesperus' and 'Phosphorus' that appear in (1) and (2) co-refer to Venus, and it is customary to think of *referential* uses of names as contributing their referent, if any, to the proposition expressed by sentences in which they occur<sup>1</sup>. As a result, the propositions expressed by the prejacent of (1) and (2) are the same, and so it remains a mystery how the reports can differ in truth-value when they are characterizations of the Greeks' beliefs by means of one and the same proposition. In other words, we expect to preserve the truth-value of (1) when substituting 'Phosphorus' for 'Hesperus', given that they make the same contribution to "what the reports say" that the Greeks believed. This is the substitution failure between co-referentials that Frege's puzzle exhibits and which we seek to account for.

Standard solutions to the substitution failure come with the rejection of some aspect or other of the overall picture about belief and/or names sketched. Thus, solutions within the Fregean camp construe 'believe' as an operator on contents individuated more finely than propositions (e.g. senses, characters, open propositions, etc.) so that we can distinguish between the beliefs ascribed by (1) and (2), respectively [3–5]. On their part, solutions within the Russellian camp claim that our truth-differing intuitions track contents that are only pragmatically associated with the reports: contents whose complete specification mentions the particular "notion", or "guise", that the speaker tacitly refers to and via which the subject is said to "agree" with the proposition *that Venus shines in the evening* [6–8]. But there are issues with the motivation and psychological plausibility of these solutions, issues which suggest that departures from the original picture should be avoided if possible. What I propose is to frame an account of our truth-differing intuitions regarding pairs of reports like (1) and (2) within a general approach to paradoxes of identity while leaving the original picture about belief and names intact.

### 1.1 Paradoxes from the Split of Individuals Through Time

The kind of account I propose can be motivated as follows. Our standard semantic theory carries a minimal metaphysics about the entities that make our statements true. This semantic theory says, again, that the contribution that a proper name makes to the

<sup>1</sup> This "referentialist intuition" is stated by the claim that proper names are *directly referential* expressions, in Kaplan's sense: i.e. expressions "[...] whose referent, once determined [...] is taken as being the propositional component" [2, p. 493].

proposition expressed is its bearer, or so in their apparently referential, literal uses. It also says that temporal adverbs (e.g. ‘tomorrow’) and attitude verbs (e.g. ‘believe’) are operators on propositions, and so that ‘Aristotle will come to dinner tomorrow’ is true if, and only if, Aristotle comes to dinner the day after the utterance, and that ‘John believes that Aristotle was a philosopher’ is true if, and only if, Aristotle was a philosopher in each of John’s doxastic alternatives. Thus, our semantic theory says that the truth of these sentences depends on how things *will* be, or *might* be with Aristotle. So if these sentences have any chance to come out true, Aristotle must be the kind of entity that can feature in the domain of future times and doxastic alternatives, i.e. the kind of entity that can *persist* through time and doxastic alternatives.

Now, persistence conditions *through time* give rise to familiar paradoxes about individual identity. The Paradox of the Ship of Theseus is an example of how these conditions allow for cases where we seem to have *two* ships that are nevertheless equal candidates to be *the* ship of Theseus. The story goes as follows. Imagine that the ship of Theseus, called ‘Victory’, is on display in a museum at time  $t$  and that, as the years go by, its worn planks are gradually replaced by new ones. Let  $t'$  be the time in which the ship has been fully restored and call that restored ship ‘Restoration’. Since ships persist through the gradual replacement of their parts, it seems clear that ‘Restoration’ is a new name for Victory, and so that Victory is identical to Restoration. Therefore, (3)

(3) At  $t'$ , Victory is in the museum

is true in that story. Now imagine an alternative story where Victory has never had its planks replaced, but instead was taken apart and stored in a warehouse. After many years, its parts were put back together in just the way they were originally ensemble. Let  $t'$  be the time in which the ship has been fully reassembled in the warehouse and call the reassembled ship ‘Reconstruction’. Since ships persist through their disassembling and reassembling, it seems clear that ‘Reconstruction’ is a new name for Victory, and so that Victory is identical to Reconstruction. Therefore, (3) is false in that story.

But now consider a “combined story” where Victory is subject to the two processes at once. Its planks have been replaced gradually in the museum and the replaced planks have been used to build an exact replica of the original ship in the warehouse. And suppose that at time  $t'$  the restoration and the reconstruction concluded. As a result, we have two ships at  $t'$ : the one we called ‘Restoration’ in the first simple story, which is in the museum, and the one we called ‘Reconstruction’ in the second simple story, which is in the warehouse. So while (4) is true, (5) is false:

(4) At  $t'$ , Restoration is in the museum

(5) At  $t'$ , Reconstruction is in the museum

But this is paradoxical: how can Restoration and Reconstruction be two distinct ships when they are both the result of processes on Victory that it could (temporally) survive? If, by contrary, they are the same ship, why does substitution of ‘Reconstruction’ for ‘Restoration’ fail to preserve truth-value? And what is the truth-value of (3) in this combined story?

## 1.2 Paradoxes from the Split of Individuals Through Possible Worlds

Not only persistence conditions through time give rise to paradoxes of identity, though. So do persistence conditions *through possible worlds* taken as doxastic alternatives. The famous Hesperus/Phosphorus-case is, I think, an example of this. Imagine that the Greeks see Venus in the evening and come to believe that there is a heavenly body  $x$  which *only* shines in the evening. Let us call that body inhabiting their doxastic alternatives ‘Hesperus’, as they themselves did. Let us follow Stalnaker [9, p. 140, 164] in that we can identify an object  $o$  in a subject’s doxastic alternative with an object  $o'$  in the actual world whenever  $o'$  plays a causal role in the generation of the subject’s beliefs about  $o'$  and properties of  $o$  tend to co-vary with properties of  $o'$ . Since Venus played a causal role in the generation of the Greek’s beliefs about  $x$  and properties of  $x$  tend to co-vary with properties of Venus (e.g. had Venus been bigger and square,  $x$  would have been bigger and square in each of the Greek’s doxastic alternatives), we may identify  $x$  with Venus and so conclude that Venus “persists” in the Greek’s doxastic alternatives. As a result, (6)

(6) The Greeks believed that Venus shines in the evening

is true in our first simple story. Now imagine instead that the Greeks see Venus in the morning and come to believe that there is heavenly body  $y$  which *only* shines in the morning. Let’s call that body inhabiting their doxastic alternatives ‘Phosphorus’, as they themselves did. Since Venus played a causal role in the generation of the Greek’s beliefs about  $y$  and properties of  $y$  tend to co-vary with properties of Venus, we may identify  $y$  with Venus and so conclude that Venus “persists” in the Greek’s doxastic alternatives. As a result, (6) is false in this second simple story.

But now let’s combine the two simple stories. The Greeks see Venus in the evening and come to believe that there is a heavenly body  $x$  which only shines in the evening. They also see it in the morning and come to believe that there is a heavenly body  $y$  which only shines in the morning. As a result, we have two heavenly bodies in the Greek’s doxastic alternatives: the one we called ‘Hesperus’ in the first simple story, which shines in the evening, and the one we called ‘Phosphorus’ in the second simple story, which shines in the morning. So while (1) is true, (2) is false. But this is paradoxical: how can Hesperus and Phosphorus be two distinct heavenly bodies when they bear the same sort of causal dependency with one and the same entity? If, by contrary, they are identical, why does substitution of ‘Phosphorus’ for ‘Hesperus’ fail to preserve truth-value? And what is the truth-value of (6) in the combined story?

It is, I think, issues concerning the (re-)identification of individuals in situations where their persistence conditions allow for a “split” that give rise to substitution failures between co-referential terms within attitude reports and our mixed intuitions about problematic reports like (6). My hypothesis is that once we get clear on how an individual can become two in a given doxastic alternative we will be able to account for how two co-referential names can become names for two distinct individuals and so fail to be interchangeable.

## 2 Substitution Failures in Quantifying-in Reports

The present approach to Frege’s puzzle is not based simply on commonalities between substitution failures in attitude reports and substitution failures in paradoxes of identity. The traditional takes on the puzzle have problems on their own that suggest the kind of approach I am advocating. For instance, Fregeans accommodate our truth-differing intuitions regarding (1) and (2) by claiming that belief contexts are “opaque”, meaning that names occurring in these contexts are used to more than simply denote their referents. A “mode of presentation” of Venus is being referred to as well. This semantic thesis is accompanied by another claim within the philosophy of mind: namely, that a belief about Venus is not fully specified without the mentioning of the way in which Venus is presented to or thought about by the subject.

Despite the role that modes of presentation are assigned by a Fregean semantics for belief, it was always kept in mind that we seem to report beliefs by quantifying into variables occupying a purely objectual position: i.e. we seem to report beliefs “*de re*”. As Quine [10, p. 178] pointed out, (7)

(7) Ralph believes that someone is a spy

has two readings, one apparently equivalent to ( $7_{\text{narrow}}$ ), which expresses Ralph’s humdrum belief that there are spies, and another apparently equivalent to ( $7_{\text{wide}}$ ), the one which belief-relates Ralph with a particular (although unspecified) individual:

( $7_{\text{narrow}}$ ) Ralph believes that  $\exists x$ . x is a spy

( $7_{\text{wide}}$ )  $\exists x$ . Ralph believes that x is a spy

But the kind of quantification that ( $7_{\text{wide}}$ ) exhibits makes the mode in which the unspecified individual is presented irrelevant to the question whether Ralph has the ascribed belief. Thus, if the Fregean account of the substitution failure exhibited in (1) and (2) in terms of a difference in the modes of presentation contributed by ‘Hesperus’ and ‘Phosphorus’ to the contents of the beliefs ascribed is correct, we should not find substitution failures in reports with quantification in, since modes of presentation do not feature in the contents of the beliefs they ascribe.

However, there are truth-differing intuitions regarding reports that, like (8) and (9)

(8) There is a heavenly body<sub>1</sub> which the Greeks called ‘Hesperus’.

The Greeks believed that it<sub>1</sub> shines in the evening.

(9) There is a heavenly body<sub>2</sub> which the Greeks called ‘Phosphorus’.

The Greeks believed that it<sub>2</sub> shines in the evening.

exhibit quantification in. This might come as a surprise, as it is commonly held that substitution between names occupying “transparent contexts” is unproblematic [10, p. 181], [11, p. 242], [12, p. 329]. But the intuitions, I believe, are there, and as Pickel [13, p. 348] claims, “[i]f one takes Frege’s puzzle cases seriously for proper names, then one should also take them seriously in cases of quantifying in”.

### 2.1 Fallacies of Intelligibility

Of course, Fregeans may react to this by claiming that on their analysis of (8) and (9) modes of presentation *do* feature in the contents of the beliefs ascribed. After all,

Fregeans charged purely objectual quantification into belief contexts of *unintelligibility*. The unintelligibility claim was given an intuitive flavor by stories like Ralph's. According to the story, "[t]here is a certain man in a brown hat whom Ralph has glimpsed several times under questionable circumstances [...] Ralph suspects he is a spy. Also there is a gray-haired man, vaguely known to Ralph as rather a pillar of the community, whom Ralph is not aware of having seen once at the beach. Now Ralph does not know it, but the men are one and the same. Can we say of this man (Bernard J. Ortcutt, to give him a name) that Ralph believes him to be a spy?" [10, p. 179] The story was meant to stress a certain conceptual defectiveness in saying that Ortcutt satisfies (or fails to satisfy) the open formula  $\lambda x$ . Ralph believes that  $x$  is a spy *independently of any way of specifying him*, and so Fregeans proposed to analyze quantifying-in reports as indirect characterizations of "notional", or "sentential" beliefs meeting an external condition [14]. On a sententialist version of the analysis, (7) would become equivalent to  $(\mathcal{T}_{\text{not}^*})$ ,

$(\mathcal{T}_{\text{not}^*}) \exists x. \exists n. n$  refers to  $x$  & Ralph believes-true the sentence 'n is a spy'

and come out true of Ortcutt and the expression 'that man on a brown hat' (as used to demonstrate Ortcutt). If now we allow that the contexts in which (8) and (9) are uttered restrict the domain quantification over modes of presentation to the class of modes of presentation of men on a brown hat, Fregeans can account for the difference in the truth-value of (8) and (9) while preserving a central role to modes of presentation in the characterization of the beliefs ascribed.

Stories like Ralph's succeed in bringing to light that something is faulty with *de re* reports like 'Ralph believes of Ortcutt that he is a spy' traditionally construed. But the Fregean analysis of reports with quantification in is motivated only if that faultiness is due to the fact that modes of presenting Ortcutt are missing in the specification of the belief ascribed. And there is reason to think that this is not the source of what is faulty with the report. Shaw [15] convincingly argues that parallel arguments for the unintelligibility of objectual quantification into temporal (and modal) constructions are fallacious. Just as it happened with (7), (10) exhibits what we would initially treat as scope ambiguity (where  $\mathbf{F}$  is an operator shifting the utterance time to any future time),

(10) The oldest senator will greet you on the Capitol steps

(10<sub>n</sub>)  $\mathbf{F} [\exists x. x$  is the oldest senator &  $x$  greets you on the Capitol steps]

(10<sub>w</sub>)  $\exists x. x =$  the oldest senator &  $\mathbf{F} [x$  greets you on the Capitol step]

and so there is quantification into temporal contexts. And just as it happened in belief contexts, substituting 'the winner of the 2006 NJ senatorial election' for 'the oldest senator from NJ' for in (11)

(11)  $\mathbf{F}$  [The oldest senator from NJ will greet you in the Capitol steps]

can affect its truth-value: although both descriptions are satisfied by Menendez at the present time, it might be that at no future time the oldest senator is the winner of the 2006 senatorial election and that only the winner will greet you. So the fact in (10<sub>w</sub>) there is no mode of presenting Menendez denoted in the statement of how things will be with him should make (10<sub>w</sub>) somewhat faulty. But, as Shaw observes, no-one would take issue with the intelligibility of (10<sub>w</sub>) on these grounds. We would deem it unqualifiedly

true whenever we can uniquely identify Menendez in some future time and see that he greets you in the Capitol steps then.

Interestingly, Shaw [15, p. 54] observes that “[w]hat helps us make sense of the *de re* reading of the description in [(10)], despite requiring quantification across a temporal operator, is that we seem to have [...] a good sense of what it is for some future person on the Capitol steps to be identical with Menendez, the man we presently identify as the oldest senator from New Jersey.” This observation, I think, helps us see what underlies the faultiness in saying of Ortcutt that Ralph’s believes him to be a spy. Arguably, Ralph has an inconsistent conception of Ortcutt: he believes that he spies and that he is not a spy. So it is reasonable to model these beliefs by means of doxastic alternatives with *two* existential instantiations of Ortcutt: one that is a spy and another that isn’t. But since there are two Ortcutts in of Ralph’s doxastic alternatives, we lack a good sense of what it is for something to be (uniquely) Ortcutt in these doxastic alternatives. Arguably, this affects the intelligibility of our reporting Ralph’s beliefs about Ortcutt. So although Fregeans are right that stories like Ralph’s put pressure on a conception of belief that somewhat allows for a single individual to satisfy both  $\lambda x$ . Ralph believes that  $x$  is a spy and  $\lambda x$ . Ralph believes that  $x$  is not a spy, the pressure vanishes once Ralph is regarded as a partially inconsistent believer, and the unintelligibility of problematic *de re* reports can be seen as simply the result of the ambiguity that some descriptions of inconsistent beliefs fall prey to.

To sum up, substitution failures in reports with quantification in cast doubts on the Fregean claim that these failures are produced by a truth-conditional import of modes of presentation. In turn, the Fregean claim that quantifying-in reports are equivalent to quantificational statements over contextually restricted classes of modes of presentation is based on a suspect diagnosis of the faultiness of some *de re* reports standardly construed. That faultiness, I have argued, has more to do with our incapacity to identify the relevant individuals across the subject’s doxastic alternatives, a capacity which is hindered when the subject has an inconsistent conception of them. All this, I think, raises a question whether we really need to introduce modes of presentation into the contents of (1) and (2) so as to account for our truth-differing intuitions, or it is other features of the reports that make us distinguish between the beliefs ascribed.

### 3 Referring to Particular Notions

The other traditional approach to Frege’s puzzle is the Russellian approach. Their advocates hold that belief should be analyzed in terms of its (Russellian, structured) proposition and its *vehicles of representation*, whether the “notions” in virtue of which the belief has the content that it has [8], the “guises” via which subjects grasp the relevant propositions [6], or sentence-like entities in one’s “belief box” [16]. On this characterization of beliefs *qua* concrete cognitive structures, it is tempting to think that the Greeks had a belief with the proposition *that Venus shines in the evening* and a ‘Hesperus’-notion yet lack a belief with that very content and a ‘Phosphorus’-notion. Thus, *if* belief reports characterized a subject’s beliefs by specifying both their contents and the notions they involved we could explain our truth-differing intuitions regarding (1) and (2).

This is, in fact, what most Russellians claim happens at a pragmatic level. Thus, advocates of the “hidden-indexical” version of the view state that “[i]n reporting beliefs, we

quite often are talking about such notions, although our belief reports do not explicitly mention them. The general solution to the puzzles is to allow a condition on particular beliefs, over and above a content condition, to be part of the claim made. [...] a specification of the notions that are supposed to be involved.” [9, p. 697] According to this version of the view, the notions the report is about need not be denoted by any expression in the report (as it happens in notionally explicit reports of the form ‘A believes that *p* via the notion *N*’), but may be provided by the speaker’s “tacitly referring” to them. Furthermore, the view takes it that the use of a particular expression usually gives the audience clues about which notion the speaker is providing. This way, the fact that the speaker uses ‘Hesperus’ when uttering (1) and ‘Phosphorus’ when uttering (2) suggests that she is referring to different notions each time.

### 3.1 Psychological Plausibility

So the Russellian hypothesis about what causes the impression that substitution between co-referential names fails is that when we deem (1) true and (2) false we are not judging the notionless proposition they semantically express (to wit, the proposition *that there is some (type of) notion/guise X such that A believes that p via X*), but rather the notion-involving enrichments explicit in reports like (12) and (13):

(12) The Greeks believed that Venus shines in the evening via a ‘H<sub>esperus</sub>’-notion.

(13) The Greek believed that Venus shines in the evening via a ‘P<sub>hosphorus</sub>’-notion

However, several authors, otherwise Russellian-friendly, have taken issue with the idea that notions are inserted into the contents of our reports by tacit reference. Thus, Braun [17, p. 560] finds it questionable “[t]hat ordinary speakers have such sophisticated thoughts and intentions about mental representations when they utter belief sentences”. In a similar vein, Schiffer [18, p. 512] claims that “[o]ne may reasonably doubt that belief ascribers mean what the hidden indexical theory requires them to mean when they ascribe beliefs”. I think there is indeed a worry about the psychological plausibility of the Russellian hypothesis. Typically, ascribers are not concerned with notions when reporting beliefs. They do not have to: a speaker may truly say that so-and-so believes that Fido is a dog “[e]ven though she is not in a position to refer to any particular mode of presentation so-and-so has for either Fido or doghood” [19, p. 100, fn. 11].<sup>2</sup> But even if ascribers did refer to notions occasionally, there are substitution failures in reports with quantification in, i.e. reports with no specification of notions or ways of grasping propositions. And this, I think, undermines much of the motivation for the claim that our truth-differing intuitions regarding (1) and (2) track notion-involving enrichments of their otherwise *conceptually* complete contents.

## 4 Back to the Hypothesis: The Paradox of Theseus Revisited

The discussion so far suggests that an account of our truth-differing intuitions regarding (1) and (2) should not assume that the speaker makes reference to anything other than particular heavenly bodies. Frege’s puzzle is a puzzle involving reports that describe an agent’s beliefs by means of a *singular proposition* about Venus. But now the question is

<sup>2</sup> See also [12, p. 328–329], [20, p. 230] or [21, p. 33] for related claims.



how (1) and (2) can differ in truth-value given that ‘Hesperus’ and ‘Phosphorus’ co-refer? My proposal is to address this question in the same way we would address the question about how (4) and (5) differ in the truth-value when evaluated at Victory’s combined story. So let me briefly sketch my preferred answer to this latter question and then show how a suitably related answer helps with Frege’s puzzle.

### 4.1 Illusions of Identity

We know that in the first simple story about the ship of Theseus,  $s_1$ , a ship has been built and named ‘Victory’, that it has been fully restored gradually, left in a museum, and that ships persist through the gradual replacement of their planks. If someone comes and names the restored ship ‘Restoration’, she will have introduced a new name for Victory. Similarly, we know that in the second simple story,  $s_2$ , Victory has been built and named ‘Victory’, that it has been disassembled and reassembled in a warehouse while keeping their planks as they were originally arranged, and that ships persist through their disassembling and reassembling. If someone comes and names the reassembled ship ‘Reconstruction’, she will have introduced a new name for Victory. In the combined story,  $s_3$ , though, Victory has been built and named ‘Victory’. However, it is fully restored gradually and left in the museum, and the planks taken off have been reassembled in a warehouse so as to make a replica of the original ship. The restored ship is named ‘Restoration’ and the reconstructed ship ‘Reconstruction’.

Now, we know that the restored ship in  $s_3$  is not identical to the reconstructed ship in  $s_3$ , if only because the former is in a museum and the latter in a warehouse. Therefore, *one of the two ships is not to be identified with Victory*. There are two options at this point: either (i) ships do not persist through gradual replacement of their planks if at the same time they are subject to their disassembling and reassembling, or (ii) ships do not persist through their disassembling and reassembling if at the same time they are subject to a gradual replacement of their planks. If (i) then (3) is true at  $s_3$ , but if (ii) then (3) is false at  $s_3$ . It is reasonable to think that our mixed intuitions about (3) when evaluated at  $s_3$  come from doubts concerning which option is right.

But let us suppose that (ii) holds: i.e. that preservation of overall form takes priority over preservation of matter. On this assumption, the names ‘Venus’ and ‘Restoration’ that originate in  $s_3$  are names for Victory, just as their respective orthographical counterparts in  $s_1$  and  $s_2$ . However, the name ‘Reconstruction’ that originated in  $s_3$  is a name for another ship, call it ‘New’, the one that has been generated in the process of reassembling the planks that used to be Victory’s.<sup>3</sup> This way we can explain the difference in truth-value at  $s_3$  between (4) and (5) in terms of a difference in the propositions that they express

<sup>3</sup> We can represent the different contents that each of these orthographically identical names have on each of these stories with the following two-dimensional concepts,

	‘Victory’			‘Restoration’			‘Reconstruction’		
	S1	S2	S3	S1	S2	S3	S1	S2	S3
s1	Victory	Victory	Victory	Victory	Victory	Victory	----	----	---
s2	Victory	Victory	Victory	----	----	----	Victory	Victory	Victory
s3	Victory	Victory	Victory	Victory	Victory	Victory	New	New	New

where the stories featuring in the vertical axis play the role of utterance situations and the stories on the horizontal axis the role of truth-supporting circumstances.

when used in  $s_3$ , i.e. by simply interpreting (4) and (5) as containing occurrences of the very names *that were originated in  $s_3$* . After all, when we considered  $s_3$  we looked at it as if it were the actual utterance situation, and so it was very natural to evaluate (4) and (5) as if the names that occurred in them had their semantic properties determined by facts in  $s_3$ . Any failure of substitution between co-referential names as well as any appearance that Victory has two “existential instantiations” in  $s_3$ , as Pickup [22] has recently claimed, is an illusion. If (ii) is the right option, the illusion is induced by the following plausible yet ultimately false conditional: (if Victory has its planks disassembled and reassembled at  $s_3$ , then name ‘Reconstruction’ that originates in  $s_3$  is a name of Victory)  $\rightarrow$  (if Victory has its planks disassembled and reassembled at  $s_3$  and its planks replaced in  $s_3$ , the name ‘Reconstruction’ that originates in  $s_3$  is a name of Victory). Its falsehood is not surprising, though. In the end, the facts that obtain in  $s_3$  are not simply the sum of those obtained in  $s_1$  and  $s_2$ : some are new (e.g. New is generated), and some of the facts that obtained in  $s_2$  are lost (e.g. Victory has been reassembled or disassembled). So we may expect that the referential facts holding in the simple stories do not “transfer” to a story that combines them.<sup>4</sup>

## 5 Extending the Approach to Frege’s Puzzle

Let us now address Frege’s puzzle as if it involved just another paradox of identity. We know that the evening body in the Greek’s doxastic alternatives is not identical to the morning body, if only because the former shines only in the evening and the latter shines only the morning. Therefore, one of the heavenly bodies inhabiting these doxastic alternatives is not identical to Venus. There are two exclusive options at this point: either (i) the evening body is Venus, or (ii) the morning body is Venus. If (i) then (6) is true; if (ii) then (6) is false. It is reasonable to think that our mixed intuitions about (6) come from doubts concerning which option is right.

Unlike the case of Victory, though, we may assume that the world is itself indeterminate as to which of the two bodies is Venus. With Victory we had two distinct criteria for identification, that of preservation of overall form and that of preservation of matter, and it was reasonable to assume that the world settles the matter about which criterion should take priority over the other. And although there is metaphysical question about how the world settles the matter, any appearance of indeterminacy about which ship is Victory in  $s_3$  is merely epistemic. But in the present case we only have one criterion of identity: that which says, roughly, that if an entity  $o$  in a possible world tends to carry information about  $o'$  by having properties of  $o'$ ,  $o$  can be identified with  $o'$ . And we have that both the evening and the morning bodies in the Greek’s doxastic alternatives tend to carry information about Venus by having properties of Venus (and, let’s assume, in the same degree). So here we are in a situation where two entities have equal right to be identified with Venus according to one and the same criteria, and so where the world itself does not seem to settle the matter.

<sup>4</sup> If instead it is (i) that holds, there would be a difference in how we should resolve the paradox at the metaphysical level, but not in the way we would account for the difference in truth-value between (4) and (5), which is my main purpose.

## 5.1 Representing Indeterminate Belief States

The moral of this is that when we represent the Greek's belief state in a possible world framework we need a representation of *indeterminate* belief states. Such a representation, I think, requires that the Greek's doxastic alternatives be modeled by two disjoint world sets,  $W_1$  and  $W_2$ . For all worlds in  $W_1$ , Venus is the evening body, whereas for all worlds in  $W_2$ , it is the morning body. Intuitively,  $W_1$  is the set of worlds that would be compatible with the Greeks' beliefs had the world settled that Venus is the evening body, and  $W_2$  the set of worlds that would be compatible with the Greeks' beliefs had the world settled that Venus is the morning body.

Aside from that, we need to stipulate that for each world  $w$  in  $W_1$  there is one (and only one) world in  $W_2$  which is *qualitatively indiscernible* from  $w$ : i.e. that it cannot be told apart from  $w$  by the way it looks. This is so because the indeterminacy hypothesis as to the identity of Venus in these doxastic alternatives translates into an irresolvable matter as to which of the two world sets in fact represents the Greek's belief state, *not* into an irresolvable matter as to whether the Greeks believe, e.g. that the heavenly body that looks such-and-such, whichever that is, has such-and-such property. So let  $f$  be a function from worlds in  $W_1$  (and  $W_2$ ) to their indiscernible worlds in  $W_2$  (and  $W_1$ ). And let  $w_1$  be a world in  $W_1$ . By assumption, Venus is the evening body in  $w_1$ . Now suppose that Venus looks red in  $w_1$ . It follows that the evening body in  $f(w_1)$  (which is not Venus) will look red in  $f(w_1)$  just like Venus does in  $w_1$ .

## 5.2 Patterns of Update on the Context

Our representation of the indeterminacy of the Greek's belief state imposes two patterns on the updating of *our* beliefs about what the Greek's believe. Let  $W_1^*$  and  $W_2^*$  be the two disjoint world sets that, for all we (ascribers) mutually presuppose, might be the Greek's actual doxastic alternatives.<sup>5</sup> If we are totally right in our presuppositions about what the Greeks believe,  $W_1^* = W_1$  and  $W_2^* = W_2$ . This is, however, unlikely, and the goal of attributing beliefs in a conversation is, we may assume, to turn  $W_1^*$  into  $W_1$  and  $W_2^*$  into  $W_2$ . The first of these patterns is that a belief attribution updates  $W_1^*$  and  $W_2^*$  by ruling pairs of world sets related by our indiscernibility function  $f$ . Thus, if a belief attribution like 'The Greeks believed that Venus doesn't look red' *directly* updates  $W_1^*$  (to wit, by ruling out those worlds in  $W_1^*$  where Venus looks red), it *indirectly* updates  $W_2^*$  by ruling their respective qualitatively indiscernible worlds where the evening body inhabiting them looks red.

<sup>5</sup> The union set  $W_1^* \cup W_2^*$  is the "derive context" set for the Greek's beliefs, as defined by Stalnaker [9, p. 157]. The derived context (for the Greeks) is determined by the basic context (i.e. the set of possible ways the world might be compatible with what all conversational partners mutually presuppose) in the following way: for each possible world in the basic context, the Greeks are in a belief state defined by the set of possible worlds compatible with what they believe in that world. The union of all these possible belief states is the set of all possible worlds that might, for all it is presupposed in the conversation, be compatible with the Greek's beliefs. On another note, this set will be typically updated when explicit belief attributions are made, and the update will proceed by the usual intersection operation: e.g. an utterance of 'A believes that S' will intersect the derive context for A's beliefs with the proposition expressed by S (as used in the utterance context).

The second of the patterns is this. A belief attribution either directly updates  $W_1^*$  or directly updates  $W_2^*$ , yet never both at once. The rationale for this pattern is the following. The names ‘Venus’, ‘Hesperus’ and ‘Phosphorus’ are all names for Venus, since their semantic properties are determined by facts obtaining in the actual world, *not* by those that obtain in the worlds as the Greeks took them to be.<sup>6</sup> Suppose that our utterance of (6) had the following effect on our derive context: it rules out from  $W_1^*$  and  $W_2^*$  those worlds where Venus does not shine in the evening. The utterance would produce two indirect updates: one that rules out those worlds in  $W_2^*$  where their evening body does not shine in the evening and another that rules out those worlds in  $W_1^*$  where their morning body does not shine in the evening. As a consequence, my utterance of (6) will count as an assertion about *both* the evening and the morning bodies in each of the doxastic alternatives. But this is intuitively incorrect, as the utterance was intended to describe how things are with only one of them in each of the doxastic alternatives. This result is blocked if our utterance of (6) is interpreted as directly updating only one of the world sets in the derive context.

Now, the second pattern of update raises an important question. Suppose that someone utters (6). Is she intending to directly update  $W_1^*$  or  $W_2^*$ ? In other words, is she making a claim about the Greek’s beliefs about the evening or the morning body? This is important, as the utterance will be true only if it directly updates  $W_1^*$ , since only then it counts as a proposal to drop all worlds in  $W_1^*$  and  $W_2^*$  where their evening bodies do not shine in the evening *and both  $W_1$  and  $W_2$  are sets of worlds where these bodies do shine in the evening*. Our earlier concern about how to decide between options (i) and (ii) shows up again and stays on the way of an unambiguous interpretation of the report. This contrasts the case of (1) or (2), reports which manage to suggest which set is being intended to directly update, and so how we should resolve the identification of Venus for the purpose of the conversation. The speaker who uses ‘Hesperus’ as in (1) suggests that we identify Venus with the evening body in each of the doxastic alternatives, most likely by exploiting the fact that this is the body which the Greeks called ‘Hesperus’, and so that  $W_1$  be taken as the set to be directly updated. If the suggestion is accepted, the utterance will drop from  $W_1^*$  those worlds where Venus does not shine in the evening, and indirectly drop from  $W_2^*$  those worlds where their evening body (not Venus) does not shine in the evening (nothing has been said about the morning bodies in  $W_1$  or Venus in  $W_2$ ). This accounts for our ordinary intuitions about the truth of (1). Likewise, the use of ‘Phosphorus’ in (2) suggests that we identify Venus with the individual that the Greeks called ‘Phosphorus’, and so that one takes  $W_2^*$  to be the set to be directly updated. If the suggestion is accepted, the utterance will drop from  $W_2^*$  those worlds where Venus does not shine in the evening, and indirectly drop from  $W_1^*$  those where the morning body does not shine in the evening. This accounts for our ordinary intuitions about the falsity of (2).

The choice of the name ‘Venus’ when reporting as in (6) does not carry a suggestion about how to identify its referent as readily as ‘Hesperus’ or ‘Phosphorus’ do, however. So unless alternative conventions are in place the report won’t proceed in accordance

<sup>6</sup> I am here departing from those views like Stalnaker’s [9], Cumming’s [3] or Pickel’s [13], to mention a few, according to which ‘believe’ shifts the contextual parameters relevant for the interpretation of the names that occur within its scope.

with the second of the patterns and so will have an unambiguous interpretation. So this is just another case where contextual resolution is required to further specify what is said (in this case, about what others believe) even after superficial context-dependency has been fully resolved.

## 6 Conclusion

I have explained how utterances of (1) and (2) differ in truth-value by exploiting the different ways we have to identify Venus across doxastic alternatives, and so by appealing to the different worlds sets that can be identified with the (coarse-grained) proposition *that Venus shines in the evening*. This is a merit of the account, as it is congenial with a simple view about quantification into belief contexts. Moreover, it is economical, as it does not resort to claims about speakers referring to representational entities (e.g. modes of presentation, notions, sentences in the mind, etc.) when reporting beliefs in order to vindicate our truth-differing intuitions. Furthermore, it is fit to account for substitution failures exhibited in reports with quantification in.

## References

1. Lewis, D.: *On the Plurality of Worlds*. Wiley-Blackwell, Hoboken (1986)
2. Kaplan, D.: Demonstratives. In: Almog, J., Perry, J., Wettstein, H. (eds.) *Themes from Kaplan*, pp. 481–563. Oxford University Press, Oxford (1989)
3. Frege, G.: Über Sinn und Bedeutung. *Philos. Rev.* **1**, 572 (1892)
4. Cumming, S.: Variabilism. *Philos. Rev.* **117**(4), 525–554 (2008)
5. Chalmers, D.: Propositions and attitude ascriptions: a Fregean account. *Noûs* **45**(4), 595–639 (2011)
6. Salmon, N.: *Frege's Puzzle*. MIT Press, Cambridge (1986)
7. Soames, S.: Direct reference, propositional attitudes, and semantic content. *Philos. Top.* **15**(1), 47–87 (1987)
8. Crimmins, M., Perry, J.: The prince and the phone booth: reporting belief reports. *J. Philos.* **86**(12), 685–711 (1989)
9. Stalnaker, R.: *Context and Content*, 1st edn. Oxford University Press, Oxford (1999)
10. Quine, W.: Quantifiers and propositional attitudes. *J. Philos.* **53**, 177–187 (1956)
11. Kripke, S.: A puzzle about belief. In: Margalit, A. (ed.) *Meaning and Use*, pp. 239–283. Reidel (1979)
12. Recanati, F.: *Direct Reference. From Language to Thought*. Blackwell, Hoboken (1993)
13. Pickel, B.: Variable and attitudes. *Noûs* **49**(2), 333–356 (2015)
14. Kaplan, D.: Quantifying in. *Synthese* **19**, 178–214 (1969)
15. Shaw, J.: De Re belief and cumming's puzzle. *Anal. Philos.* **56**(1), 45–74 (2015)
16. Richard, M.: How I say what you think. *Midwest Stud. Philos.* **14**, 317–337 (1989)
17. Braun, D.: Understanding belief reports. *Philos. Rev.* **107**(4), 555–595 (1998)
18. Schiffer, S.: Belief ascription. *J. Philos.* **89**(10), 499–521 (1992)
19. Schiffer, S.: Belief ascription and a paradox of meaning. *Philos. Issues* **3**, 89–121 (1993)
20. Bach, K.: Do belief reports report beliefs? *Pac. Philos. Q.* **78**, 215–241 (1997)
21. Hanks, P.: Structured propositions as types. *Mind* **120**(477), 11–52 (2011)
22. Pickup, M.: A situationalist solution to the ship of theseus puzzle. *Erkenntnis* **81**, 973–992 (2015)