

# The Puzzle of Names

EDUARDO GARCÍA-RAMÍREZ

*Universidad Nacional Autónoma de México*

| 69

**Abstract:** The literature is unclear on whether informativeness and substitution failure present two different problems to a theory of proper names for natural languages. In this paper I argue that the problems are intimately related and show how the tradition has failed to notice it. I offer an account of why this is so and present a new version of the problem. This finding is directly relevant for the prospect of discovering an adequate solution to the problems in question. The resulting formulation gives us a good idea of the kind of solution required.

**Key-words:** substitution failure, informativeness, natural language.

**Resumen:** La literatura en torno a los problemas de informatividad y fallas de sustitución no es clara respecto a si se trata de dos problemas independientes que deba resolver toda teoría satisfactoria de los nombres propios del lenguaje natural. En este texto argumento que

los problemas están íntimamente relacionados y muestro cómo la tradición no se ha percatado de esto. Presento una explicación de por qué ha sucedido esto y ofrezco una versión novedosa del problema. Este descubrimiento es directamente relevante para todo intento por encontrar una solución satisfactoria del problema en cuestión. La formulación resultante nos da una buena idea sobre el tipo de solución requerida.

**Palabras clave:** fallas de sustitución, informatividad, lenguaje natural.

## 1. Introduction

Frege (1892) presents the problem of informativeness as a problem about some uses of identity statements. He does not present it as a problem of belief ascription or propositional attitudes. Russell (1905) presents the problem of substitution failure as a problem about belief ascription or propositional attitudes. He does not present it as a problem of informativeness. This way of posing the problems has proven fruitful. Since Frege and Russell, many<sup>1</sup> have followed them in conceiving *informativeness* and *substitution failure* as two separate and independent issues; thus, allowing us to understand better what these problems are about.

70 |

This attitude is evident in the way in which different philosophers have tried to solve the problems. Millians seem to agree that they must deal with two different issues. Kripke (1979) addresses the problem of substitution failure, but does not say much (if anything) about informativeness. Something similar goes on in Soames (1987).<sup>2</sup> Likewise with Descriptivists. In his presentation of what he dubs “the Frege–Russell problems” Kroon (2004) takes them to be different, independent problems. Even contemporary reviews of the topic still present them as two separate issues.<sup>3</sup>

I believe we have learned enough. Informativeness and substitution failure are not two separate, independent, issues. They go hand-in-hand. They are so closely related to each other that one wonders whether they are really two different problems. In section 2 I will present each problem with some detail. Section 3 offers four different solutions from philosophers with opposing theoretical views. The evidence suggests that the alleged ‘problems’ are intimately

—

<sup>1</sup> See McKay and Nelson 2010 and Swanson 2006.

<sup>2</sup> See also Soames 2002.

<sup>3</sup> See Braun 2007 and Lycan 2006

related since *any* solution to one problem is proved to be enough to solve the other. Section 4 gives a brief description of the historical context. It seems that philosophers, in general, have failed to notice this intimate relation. In section 5 I offer an explanation, in psychological terms, of why these issues are so intimately related. I conclude in section 6 by describing such relation and reformulating the problems in a way that points towards the kind of solution required.<sup>4</sup>

## 2. The Puzzles

### *Informativeness*

Frege's (1892) puzzle about *informativeness* can be easily stated. The sentences in 1 and 2 differ, information-wise, amongst each other. For example, (1a) is trivial while (1b) is not, and (2b) may convey new information to an addressee who is already familiar with 2a. Yet, they merely differ –i.e., (1a) from (1b) and (2a) from (2b)– in their use of different but coreferential names: 'Mark Twain' and 'Samuel Clemens'.

(1a) Mark Twain is Mark Twain.

(1b) Mark Twain is Samuel Clemens.

(2a) Mark Twain wrote *Adventures of Huckleberry Finn*.

(2b) Samuel Clemens wrote *Adventures of Huckleberry Finn*.

I 71

Given a plausible view of names according to which they contribute only their referents, we have a problem explaining where this difference of informativeness comes from. The problem seems clear, but what is not is the very notion of informativeness. What does it mean to say that (1b) is informative while (1a) is not, or that (2b) can be informative in contexts where (2a) is not?

Frege's (1892) own account is not helpful. He says, for example, that (1a) is "valid a priori" and "called analytic" while (1b) provides "very valuable extensions of our knowledge" and is not "a priori". He also says that (1b), but not (1a), can be used to express a discovery and that, unlike (1a), it is not self-evidently true. These distinctions are not very useful. We know, thanks to Kripke (1980), that being

---

<sup>4</sup> Before getting started, a brief warning. I am concerned with informativeness and substitution failure as they pose conditions on a satisfactory account of proper names *for natural languages* like English and Spanish. Thus, I will not be concerned with how, or even if, these problems arise for formalized languages.

knowable *a priori* does not preclude some truth from expanding our knowledge. We are left with the unclear idea of “knowledge expansion”.

In a more recent rendition of the problem, Braun says that (1a) and (1b) –and *a fortiori* (2a) and (2b)– differ in informativeness because “a rational, competent, speaker could understand both and yet think that one is true and the other false” (Braun 2007: 492). Notice, however, that the criterion of informativeness is psychological: two sentences differ in informativeness if a normal subject can give opposing credence to both (e.g., believing one and disbelieving the other). Thus, mental states play a central role in this understanding of informativeness. Indeed, this is what most philosophers do. They either explicitly appeal to belief and other mental states, as Wettstein (1989) does, or presuppose them, as Kroon (2004) does.<sup>5</sup>

Thus, another way to understand the differences in informativeness, for example, between (2a) and (2b), is to notice that (3a) and (3b) can both be true.

(3a) Andy believes that Mark Twain wrote *Adventures of Huckleberry Finn*.

(3b) Andy does not believe that Samuel Clemens wrote *Adventures of Huckleberry Finn*.

72 |

If this is, as it seems, our best way to understand what we mean by “informativeness”, then it is important to note that the notion itself is speaker, or subject, relative. Here is an initial formulation of the problematic notion.

**INF:** For any pair of sentences *S1* and *S2*, which are not necessary falsehoods and which vary only in the use of different coreferential names, a normal subject *A*, and a context *c*, *S1* and *S2* differ in informativeness in *c* iff *A* believes one and does not believe the other in *c*.

This account has a limitation. There are many ways in which an utterance of mine can be informative. For example, it can be informative in virtue of its presuppositions or by updating the information you have about my mental states (e.g., when you already know who won the prize and I truly say: Jon won the prize). The problem of informativeness, and my account thereof, is not concerned with any of these. It focuses on the way in which an utterance can be informative in virtue of its *subject matter*. A good feature of INF is that it explains not only the

---

<sup>5</sup> See also Lycan 2006.

difference of informativeness among necessary truths, like (1a) and (1b), but also among contingent truths, like (2a) and (2b).

Two important claims follow from what I have said. First, informativeness is not a property that sentences have *simpliciter* but relative to a speaker –if you prefer we can say it is a relation.<sup>6</sup> This claim is of great importance. If it is true then a difference of informativeness is directly tied to attitudinal mental states.

Second, in order to describe how the difference of informativeness takes place, particularly between speakers and pairs of sentences, one *must* allow for the possibility of substitution failure as in (3a) and (3b). One cannot explain why (2a) and (2b) differ for Andy without accepting the truth of (3a) and (3b) and, ultimately, of (4a) and (4b), which are both instances of substitution failure of coreferential terms within belief reports.

(2a) Mark Twain wrote *Adventures of Huckleberry Finn*.

(2b) Samuel Clemens wrote *Adventures of Huckleberry Finn*.

(3a) Andy believes that Mark Twain wrote *Adventures of Huckleberry Finn*.

(3b) Andy does not believe that Samuel Clemens wrote *Adventures of Huckleberry Finn*.

(4a) Andy believes that Mark Twain is Mark Twain.

(4b) Andy does not believe that Mark Twain is Samuel Clemens.

I 73

The problem was this: (2a) and (2b) differ in informativeness, but vary simply in using different coreferential terms. And this was the cause: a normal agent can believe one without believing the other, which in turn gives place to substitution failure. It seems, then, that informativeness becomes a problem only when there is substitution failure.

### *Substitution Failure*

But, what is substitution failure? There is substitution failure whenever the substitution of coreferential terms fails to preserve the truth-value of the original sentence. The sentences in 5, all of which are true, illustrate this.

(5a) Andy believes that Mark Twain wrote *Adventures of Huckleberry Finn*.

---

<sup>6</sup> Or, if you are not satisfied, we can define it as BEING INFORMATIVE-TO-S and then claim it is a property that statements have *simpliciter*. That does not affect my argument, since it is still (now explicitly) referring to the mental states of the speaker.

(5b) Andy does not believe that Samuel Clemens wrote *Adventures of Huckleberry Finn*.

(5c) Mark Twain is Samuel Clemens.

Russell (1905) thinks that (5a)–(5c) create a problem for any theory of denoting expressions like ‘Mark Twain’ and ‘Samuel Clemens’. He starts by assuming Leibniz’s principle of Indiscernibility of Identicals (LI) according to which if Mark Twain is identical with Samuel Clemens, then whatever is true of Mark Twain is true of Samuel Clemens. Russell adds: “either may be substituted for the other in any proposition without altering the truth or falsehood of that proposition” (Russell 1905: 215). This is, briefly put, the principle of substitutivity. The principle seems intuitive and almost evidently true. That is precisely why its failure is problematic. It is so evidently true that it should not fail; the substitution of two terms that denote exactly the same object (e.g., ‘Mark Twain’ and ‘Samuel Clemens’) should not alter the truth or falsehood of the proposition expressed.

One can choose among many options to solve the problem. Two obvious ones are: (i) to claim that names contribute more than just their referent, or (ii) to directly reject the possibility of substitution failure (e.g., claiming that (5a) and (5b) cannot both be true). Descriptivists, like Frege, have followed (i) claiming that the semantic contribution of a name is not exhausted by its referent but includes a description. Millians, who believe that names semantically contribute just their referent, have followed both (i) and (ii): the former with non-semantic strategies<sup>7</sup> and the latter with different counterintuitive moves.<sup>8</sup> One of the counterintuitive results of the latter option is that normal subjects turn out to be either illogical or irrational.<sup>9</sup>

Following option (i) requires one to accept that subjects may ascribe incompatible properties to the same object. If the sentences in 5 are true, it seems safe to infer that Andy ascribes both properties *AUTHORING ADVENTURES OF HUCKLEBERRY FINN* and *NOT AUTHORING ADVENTURES OF HUCKLEBERRY FINN* to the same object, namely, Mark Twain. Accepting this is certainly not the same



<sup>7</sup> See Stalnaker 1987, Recanati 1993, and Swanson 2006.

<sup>8</sup> See Braun 2006, Salmon 1986 and 2006, Soames 2002, and plausibly Marcus 1981.

<sup>9</sup> Salmon 2006 and Braun 2006 seem to follow this path. They distinguish between illogical and irrational attitudes. Surprisingly enough Salmon does not say what illogical amounts to, see 1989 and 2006. Braun 2006 claims that a rational speaker may accept the identity statement (5c) and the substitution of names in (5a) and (5b). Doing so is rational. This entails a logical contradiction, however. Braun accepts this is illogical. He does not say what the difference is between being “illogical” and being “irrational”. Braun and Salmon have faced an intense attack from Schiffer; see Schiffer 2006 and 1987.

as ascribing irrationality to normal subjects, but it is still problematic, specially if, as is generally admitted, normal subjects with normal conceptual abilities do it. So we have a further problem: substitution failure is caused by lousy property attributions, and what explains this phenomenon? The traditional answer is to claim, explicitly or not, that normal subjects may sometimes fail to notice the truth of some identity statements. This seems especially true of cases like (1a)-(1b) and (4a)-(4b), where it is obvious that competent speakers do ignore such information.

There are, as is natural, many different stories you can tell about this ignorance –I will present two of them in this paper– but if the story is to be of any use, it must explain how acquiring knowledge about identity statements can make a cognitive difference by precluding competent speakers from attributing incompatible properties to the same object. This, in turn, is tantamount to claiming that true identity statements are *informative*. If they were not, then coming to learn them would not preclude the subject from making such lousy property attributions. And thus, we would not have an account of substitution failure.

As you can tell, this gets us right into the problem of informativeness. For claiming that true identity statements are, or can be informative, requires one to assume a difference between (1a) and (1b), trivial and informative identity statements. And this difference of informativeness between pairs of sentences that vary in using different coreferential terms is, as we saw in the previous section, a problem. This suggests a formulation of substitution failure among the following lines.

I 75

**SUBS:** For any two coreferential names ‘*N*’ and ‘*M*’, any context *c*, and a normal subject *A*: ‘*N*’ and ‘*M*’ are subject to substitution failure within reports in *c* of *A*’s first-order belief states iff *A* does not believe that ‘*N is M*’ is true in *c*.

Two important limitations apply. There are many ways in which two coreferential names may fail to be substitutable within belief reports. They may fail to substitute because the reports are of different orders (e.g., Andy believes that Samuel Clemens is Mark Twain, but he does not believe that he believes that he believes that Samuel Clemens wrote *Adventures of Huckleberry Finn*, even though he believes that Mark Twain did; he has not thought about this). Traditionally, the problem has been taken to be about first-order belief reports. The account above is consistent with this limitation.

Second: parallel to the case of informativeness, the failure to substitute *salva veritate* concerns only the subject matter of the belief report: the ascriber’s mental states. Substitution failure is not something that concerns the embedded sentence in a first order belief report. Presumably, there is no substitution failure in such

cases (e.g., ‘Mark Twain is Mark Twain’ and ‘Mark Twain is Samuel Clemens’ have the same truth-value).

Note that SUBS describes not only cases where names fail to be substituted within belief reports with a contingent truth in the embedded clause: e.g., (5a) or (5b). It also accounts for substitution failure within belief reports that have a necessary truth in the embedded clause, e.g., (4a) and (4b). This constitutes another parallel between substitution failure and the problem of informativeness. Contingent and necessary truths appear to be on the same standing with respect to both. As before, I believe the relation between both problems is close.

(4a) Andy believes that Mark Twain is Mark Twain.

(4b) Andy does not believe that Mark Twain is Samuel Clemens.

(5a) Andy believes that Mark Twain wrote *Adventures of Huckleberry Finn*.

(5b) Andy does not believe that Mark Twain wrote *Adventures of Huckleberry Finn*.

(5c) Mark Twain is Samuel Clemens.

76 | Substitution failure was a problem because (5a) and (5b) are both true even though the embedded sentences vary simply in their use of different coreferential terms. And this was the cause: a normal agent may fail to notice the truth of an identity statement such as (5c). This ignorance, however, gives place to the problem of differences of informativeness. It seems, then, that there is substitution failure only when informativeness becomes a problem.

In this and the previous section I have shown how differences of informativeness come with substitution failure and vice versa. This suggests that, in so far as they are problematic, *informativeness* and *substitution failure* go together. In section 3, I will present four different solutions to the problems. I will describe how they are meant to solve one or the other and show how they also solve the remaining problem, which they were not (at least explicitly) meant to solve. This suggests that not only the problems, but also their solutions, go hand-in-hand.

### 3. Case Studies

*Frege*

Frege (1892) explicitly solves *informativeness* by appealing “senses”. He does not explicitly offer a solution to *substitution failure*, but his views on indirect speech offer an explanation. My goal will be to show that any one of



these moves, either appealing to indirect speech or to “senses”, is good enough to solve both problems.

Frege (1892) claims that there is only one way to account for differences of informativeness:

A difference could arise only if the difference of the signs corresponds to a difference in the way in which the designated objects are given (Frege 1892: 199).

And later on:

[It] is plausible to connect with a sign (...) not only the designated object, which may be called the *nominatum* of the sign, but also the sense (connotation, meaning) of the sign in which is contained the manner and context of presentation (Frege 1892: 200).

Briefly put, proper names have referents and senses. Different proper names for the same referent may have different senses. How does this solve the problem of informativeness? Recall our problematic sentences:

I 77

(1a) Mark Twain is Mark Twain.

(1b) Mark Twain is Samuel Clemens.

If sentence (1b) is true, then ‘Mark Twain’ and ‘Samuel Clemens’ have the same referent. Yet they differ in informativeness. How can this be? Answer: very simply, the names have different senses that determine the sentences’ meanings. ‘Mark Twain’ presents Mark Twain as, say, *the F*; whereas ‘Samuel Clemens’ presents Mark Twain as, say, *the G*. Sentence (1a) conveys the information that *the F is the F*, which is trivial. Sentence (1b) conveys the information that *the F is the G*, which is informative. End of story.<sup>10</sup>

Now, let me ask if this explains substitution failure. Remember our other problematic sentences:

(5a) Andy believes that Mark Twain wrote *Adventures of Huckleberry Finn*.

---

<sup>10</sup> I do not intend to take Frege’s story as a version of Descriptivism. I simply use definite descriptions in pretty much the same way as Frege does: as a way to represent a *sense*.

(5b) Andy does not believe that Samuel Clemens wrote *Adventures of Huckleberry Finn*.

(5c) Mark Twain is Samuel Clemens.

Given the truth of (5c) and the principle of substitutivity, we should be able to substitute 'Mark Twain' for 'Samuel Clemens' in (5b) without altering its truth-value. We cannot do this because it is inconsistent with (5a) being true. Why does substitution fail? Couldn't we give the same story as above? Couldn't we say, for example, that names have different senses that determine their meanings? 'Mark Twain' presents Mark Twain as, say, *the F*; whereas 'Samuel Clemens' presents Mark Twain as, say, *the G*. Hence, the embedded sentence in (5a) conveys information about *the F*, while that in (5b) conveys information about *the G*.

But, the Fregean skeptic may reply, how could this account for substitution failure? If there is no change of reference so that attitudinal contexts, (5a) and (5b), shift their reference from an object to a sense, how can we account for the truth of both (5a) and (5b)? Senses by themselves are not enough to determine truth-values, they need to become referents to do so. Or so the Fregean would say.<sup>11</sup>

I have a short and a not so short reply to this worry. The short answer is that the way to account for the truth of (5a) and (5b) by merely appealing to different senses (i.e., without appealing to reference shifting) is inconsistent with Frege's loved principle of compositionality. The not so short reply requires that we recall the description of substitution failure offered above.

78 |

**SUBS:** For any two coreferential names '*N*' and '*M*', any context *c*, and a normal subject *A*: '*N*' and '*M*' are subject to substitution failure within reports in *c* of *A*'s first-order belief states iff *A* does not believe that '*N is M*' is true in *c*.

According to this formulation, we have an account of substitution failure as soon as we can explain why a subject's first order belief states change in the presence of different but coreferential names. We can do this by appealing to different senses. If different senses are enough to prompt different attitudes—as Frege proposes in his account of informativeness—, subjects that associate different senses to each different coreferential name may have different first-order belief states.

Of course, this explains why they have different attitudes, but the Fregean wants to now why (5a) and (5b) have the truth-value they have. How can this

—

<sup>11</sup> Thanks to an anonymous referee for pointing this out.

happen? The answer now should be clear. (5a) and (5b) are both true because they both represent an Andy's state of mind accurately, and this is ultimately so precisely in virtue of the fact that Andy associates different senses to the names 'Mark Twain' and 'Samuel Clemens'.

This is certainly not the kind of account that the Fregean is used to, for it is *not* an account of the truth-values of (5a) and (5b) are determined by their logical form and the meaning of their parts. It is *not*, that is to say, a compositional account. There are psychological features (i.e., association and processing of different senses) that get into the account of the subjects belief-states but do not get into the semantics or syntax of the attitudinal reports. Furthermore, on this account, directly derived from Frege's theory of informativeness, (5a) and (5b) may have the same truth value even if their linguistic parts make exactly the same contribution and the logical form of (5b) is the negation of that of (5a). This should also be enough to explain why Frege did *not* accept this account of substitution failure, for it amounts to admitting that natural languages are not fully compositional.

This much is clear by now: Frege's account of the differences in informativeness *also* explains why there is substitution failure (albeit in a non-compositional way). This, however, is not something Frege claims explicitly. He seems to have a separate treatment of substitution failure. It is important to note that Frege himself seems to be drawing a connection between informativeness and propositional attitudes, since the very idea of there being different senses opens the possibility of there being different attitudes in equally competent speakers.<sup>12</sup> Still, as I will show in what follows, Frege offers different accounts for both informativeness and substitution failure. It will be important to understand why he does that.

Here is what Frege says about direct and indirect speech, which turns out to solve the problem of substitution failure.

When words are used in the customary manner then what is talked about are their nominata. (...) In indirect discourse words do not have their customary nominata; they here name what customarily would be their sense. (...) The indirect nominatum of a word is therefore its customary sense (Frege 1892: 200).

The embedded sentences in (5a)-(5b) are used indirectly. Frege's theory says that the expressions used in those sentences, say, 'Mark Twain' and 'Samuel

---

<sup>12</sup> I am grateful to an anonymous referee for pointing this out.

Clemens', change their reference. They now refer to their customary senses, say, *the F* and *the G* respectively. How can we use this to explain substitution failure? Answer: very simply, the names are used indirectly, so, according to Frege, they refer to their customary senses. They have different customary senses. Hence, they refer to different things. So the principle of substitutivity does not apply. Problem solved.

Whether or not you think this solution is the same as above depends on what you take to be doing the explanatory work. Is it the fact that names change their reference in belief reports? Or is it the fact that the names in question have different senses? Here are some reasons to think it is the latter. On the one hand, if, for some reason, the names were to have the same sense, regardless of whether they change reference from nominatum to customary sense, substitutivity holds—which obviously does not solve the problem. On the other hand, if, for some reason, names were to keep their customary reference and sense in indirect speech, assuming they have different senses, substitutivity would fail because the names have different meaning. This strongly suggests that the difference in senses is doing the explanatory work. Still, the explanations seem to be different. Frege's indirect-speech solution says that substitutivity fails because the names are not coreferential within the scope of attitude verbs. The Fregean solution I gave before claims that the principle fails because the names are not synonyms, both in direct and indirect speech. Of course, one might as well put it differently and claim that both solutions say that substitutivity fails because the names have different senses.

80 |

This account, in terms of indirect speech and reference, presupposes the existence of senses. So, it presupposes Frege's account of differences of informativeness. But it is not clearly a mere extension of the latter. Some such extension is the Fregean account I offered above. Frege does more than just *extending* his view of informativeness, he adds an important claim: that attitude contexts are special in that reference gets shifted. As I showed above, one can explain substitution failure, without making such a claim, by merely adding senses. Let us suppose that this solution is different from the one about informativeness. If so, then I am obliged to ask: does the indirect-speech solution solve the problem of informativeness?

Suppose all we have is Frege's theory of indirect speech. It says that names with different senses in direct speech will not be coreferential within belief reports. This offers a simple solution to the problem of informativeness. Why do (1a) and (1b), or (2a) and (2b) differ in informativeness? Answer: we know the names 'Mark Twain' and 'Samuel Clemens' are coreferential in all of them. That is why (1a)-(1b), (2a)-(2b), have the same truth-value. Nevertheless, we also know that they are not coreferential in (5a)-(5b), because they cannot be substituted

*salva veritate*. Thus, says the indirect speech theory, the names must have different senses. If so, then ‘Mark Twain’ and ‘Samuel Clemens’ do not convey the same information. This explains why (1a)–(1b), and (2a)–(2b), differ in informativeness. Problem solved!

Frege’s account of the differences in informativeness explains *substitution failure*. Frege’s indirect-speech account of substitution failure explains the differences in informativeness. If a theory works with one problem it works with the other. Both solutions go together.

### *Recanati on belief reports*

Recanati (1993) offers an account of substitution failure based on the dual claim that declarative sentences, in general, are sensitive to embedding contexts and that ‘that’-clauses (e.g., ‘Andy believes that...’) are peculiar contexts where the content expressed by the embedded sentence is underdetermined: there is no single, semantically determined, way of finding out what is expressed by the embedded sentence. In this sense, ‘that’-clauses are ambiguous. As such, the content expressed by the embedded sentence need not coincide with the one it would express outside of the clause. This allows for the possibility that ‘Mark Twain wrote *Adventures of Huckleberry Finn*’ and ‘Samuel Clemens wrote *Adventures of Huckleberry Finn*’ may express the same content outside but not inside a ‘that’-clause.

Recanati (1993) agrees that names *semantically* contribute only their referent, but claims that referring expressions, in general, *pragmatically* contribute a way of presenting their referent. As with Frege’s “senses”, it is unclear what these “modes of presentation” are. Recanati suggests that ‘Mark Twain’, for example, presents Mark Twain as MARK TWAIN or, alternatively, as CALLED ‘MARK TWAIN’. It is this pragmatically conveyed information that helps “enrich” what is otherwise the underdetermined content of the embedded sentence. The mode of presentation, whatever that is, is added to the content of the embedded sentence. This is, claims Recanati 1993, a pragmatic process that becomes truth-conditionally relevant.

This is how the account explains substitution failure:

Suppose someone says (5a). If the hearer contextually assumes that the ascriber is being ‘faithful’ to the believer, he will be led to assume that the ascribed belief concerning Mark Twain involves a mode of presentation (...) which includes the descriptive concept ‘called “Mark Twain”’. Hence, the hearer will assume that the believer thinks of Mark Twain as ‘Mark Twain’; he will take the reference of the ‘that’- clause as a quasi-singular

proposition involving a certain (*de re*) concept over and beyond the reference of (...) ‘Mark Twain’. (Recanati 1993: 358)<sup>13</sup>

And what is a “quasi-singular” proposition? Assuming propositions are structured sets of objects (e.g., referents and properties), the proposition expressed by ‘Mark Twain wrote *Adventures of Huckleberry Finn*’ within (5a) is something like (5 $\phi$ ):

(5a) Andy believes that Mark Twain wrote *Adventures of Huckleberry Finn*.

(5 $\phi$ ) << Mark Twain, ‘Mark Twain’> AUTHORIZING *ADVENTURES OF HUCKLEBERRY FINN* >

Similarly, the proposition expressed by ‘Samuel Clemens wrote *Adventures of Huckleberry Finn*’ in (5b) is something like (5 $\psi$ ):

(5b) Andy does not believe that Samuel Clemens wrote *Adventures of Huckleberry Finn*.

(5 $\psi$ ) << Mark Twain, ‘Samuel Clemens’>, AUTHORIZING *ADVENTURES OF HUCKLEBERRY FINN* >

82 |

If (5 $\phi$ ) and (5 $\psi$ ), respectively, are the contents of the embedded sentences in (5a) and (5b), then substitution failure should be expected: using different names, ‘Mark Twain’ or ‘Samuel Clemens’, delivers different contents and, hence, belief reports with different truth-values.

Suppose you like this reply to the problem of substitution failure. If I am correct, it should also explain the differences in informativeness. Remember the puzzle? Sentences (1a)-(1b), and (2a)-(2b), differ in informativeness even though they vary merely in using different coreferential names.

(1a) Mark Twain is Mark Twain.

(1b) Mark Twain is Samuel Clemens.

(2a) Mark Twain wrote *Adventures of Huckleberry Finn*.

(2b) Samuel Clemens wrote *Adventures of Huckleberry Finn*.

The sentences in question do not vary in truth-value, so we should

---

13 I have modified the text to fit the examples of this paper. In the original text Recanati has ‘Cicero’ for ‘Mark Twain.’

not expect them to vary in terms of the contents they express. According to Recanati's "enrichment" story, embedded declarative sentences are enriched by "modes of presentation", which are themselves pragmatically conveyed by the names used. This is so because, in general, referential terms convey these "modes of presentation", but they do so in a pragmatic way. If this is so, then, all uses of 'Mark Twain', and of 'Samuel Clemens', pragmatically present their referents as 'Mark Twain', and as 'Samuel Clemens'. This happens even in declarative sentences such as (1a)-(1b) and (2a)-(2b). This is enough to explain the differences in informativeness.

Even though they express the same proposition, sentences (1a)-(1b), and (2a)-(2b), do so by pragmatically conveying *different* modes of presentation of Mark Twain. Whereas (2a), for example, presents the author of *Adventures of Huckleberry Finn* as 'Mark Twain', (2b) presents the same object as 'Samuel Clemens'. Hence, the difference in informativeness between both sentences. Similar considerations apply to (1a)-(1b) and their difference in informativeness.

It seems that Recanati's ambiguity/enrichment story on 'that'-clauses not only explains substitution failure but also the differences in informativeness. Solving one problem goes hand-in-hand with solving the other.

#### *Plantinga and the Boethian compromise*

I 83

I have presented two different solutions to the puzzles: Frege's descriptivist and Recanati's anti-descriptivist. They both confirm my hypothesis: *informativeness* and *substitution failure* are solved together. Let me now present another, quite different, descriptivist solution: Plantinga's (1978) account of substitution failure.

Plantinga (1978) claims of proper names that: (a) they express properties; (b) these properties are essential to the referent; and that (c) "different proper names of an object can express logically equivalent but epistemically inequivalent essences of that object". (a) is meant to capture Frege's intuitions while (b) satisfies the Boethian compromise.<sup>14</sup> As for (c), things are more complex.

Plantinga relies on a simple view of essential properties: *P* is an essential property of *x* iff in every world where *x* exists *x* has *P* and in no world is there an object *y*, distinct from *x*, which has *P*. According to this view, any given object has a plethora of essential properties. One can easily transform a contingent property of a given object into an essential one by turning it into a "world-indexed"

---

<sup>14</sup> According to Plantinga, Boethius was the first philosopher to claim "that names express individual essences" (see Plantinga 1978: 128). He calls this view "Boethianism".

property. In this sense *being the author of "The Boethian Compromise" in the actual world* would be an essential property of Plantinga.

Plantinga points out an inconvenience. Conceived as such, any essential property of a given object is logically equivalent with *all* the remaining essential properties of the same object. For example, in all and only those worlds in which Plantinga has the property above mentioned he also has the property of *being the author of The Nature of Necessity in the actual world* (and a good number of other properties). It follows from this (together with the principles of Plantinga's theory) that any proper name of an object expresses all the essential properties of that object.

That this is a problem is shown by the fact that it does not solve any of the puzzles. Consider the problem of informativeness: why do (1a)-(1b), or (2a)-(2b), differ?

(1a) Mark Twain is Mark Twain.

(1b) Mark Twain is Samuel Clemens.

(2a) Mark Twain wrote *Adventures of Huckleberry Finn*.

(2b) Samuel Clemens wrote *Adventures of Huckleberry Finn*.

84 |

Suppose then that we accept (a) and (b). It follows that 'Mark Twain' and 'Samuel Clemens' express the same properties of Mark Twain, namely, all of its essential properties. Hence, it follows that (1a) and (1b) express exactly the same proposition and, thus, do not differ content-wise. So we have no explanation of the difference in informativeness.

Plantinga's solution is simple: accept (c). In other words, we must accept that, even though they express logically equivalent contents, 'Mark Twain' and 'Samuel Clemens' may express epistemically inequivalent contents. Plantinga does not say much about what makes two properties be "epistemically inequivalent". He does have a very simple, and very Fregean, criterion for making such distinctions between properties. Consider (2a) and (2b).

As you may recall, someone might believe (2a) without believing (2b). Andy is, in fact, such a person. According to Plantinga, Andy's different propositional attitudes towards (2a) and (2b) are sufficient to show that (2a) and (2b) express different propositions: "and this is due to the fact that their singular terms express different properties". 'Mark Twain' and 'Samuel Clemens' both express essences of Mark Twain, "but epistemically inequivalent and hence different essences" (Plantinga 1978: 129-30).

Now, to say that Andy might believe (2a) without believing (2b) is the same as saying that (5a) and (5b) may both be true; and this is the problem of substitution failure.



(5a) Andy believes that Mark Twain wrote *Adventures of Huckleberry Finn*.

(5b) Andy does not believe that Samuel Clemens wrote *Adventures of Huckleberry Finn*.

In other words, Plantinga's reason to think that coreferential names may express epistemically inequivalent properties is the very same that makes us think that substitution may fail for some coreferential names. It is not surprising, then, that Plantinga's theory solves this problem: it is built into the theory.

Why is there substitution failure? Answer: very simply, proper names express essential properties of objects. Coreferential names may express epistemically inequivalent properties. When such is the case, sentences that differ merely in using different coreferential names may express different propositions. When this happens, there is substitution failure. End of story. This solution, however, is somehow circular. There is substitution failure because sentences using different coreferential names may express different propositions. This begs the question, how do we know they express different propositions? Plantinga's answer is, now, circular: because someone may believe what is said when using one name without believing what is said when using the other name (i.e., because substitution fails).

Plantinga also offers an account of informativeness. In his own words,

Since 'Mark Twain' and 'Samuel Clemens' express epistemically inequivalent essences of Mark Twain, (1a) and (1b) express epistemically inequivalent propositions, so that (1b) can be informative.<sup>15</sup>

I 85

Why is there a difference in informativeness between (1a) and (1b)? Answer: very simply, proper names express essential properties of objects. Coreferential names may express epistemically inequivalent properties. When such is the case, sentences that differ merely in using different coreferential names may express different propositions; one may be informative while the other is not. Hence, a difference in informativeness between (1a) and (1b), and similarly between (2a) and (2b). End of story.

As you can see, Plantinga's solution to the problem of informativeness is the same as Plantinga's solution to the problem of belief ascriptions. To make this even clearer, note Plantinga's own remarks concerning his account of informativeness: "Surely this is the natural and intuitively plausible position; surely

---

<sup>15</sup> Plantinga 1978: 134. I have modified the text to fit the examples of this paper. In the original text Plantinga has 'Hesperus' for 'Mark Twain' and 'Phosphorus' for 'Samuel Clemens'. Sentences (1a) and (1b) correspond to his (26) and (19), respectively.

a person could believe (1a) and (2a) without believing either (1b) or (2b).”<sup>16</sup> If I am not misreading Plantinga, he claims that “the natural and intuitively plausible position” about differences in informativeness is the one that explains substitution failure within attitudinal contexts.

Once again, we reach the same conclusion. Plantinga’s account of substitution failure explains the differences in informativeness. Even Boethians think that the solutions for these problems go together.

*Stalnaker on two-dimensionalism*

Stalnaker (1978) offers a theory of assertion in terms of an interaction between the content of the uttered sentence and the context of the conversation—hence the two dimensions. Context affects content in two ways: it determines what is said and its truth-value. This can be represented formally by means of a two-dimensional matrix, such as *A*, where propositions are represented in the horizontal lines.

<i>A</i>	W1	W2
W1	T	T
W2	F	F

86 |

Content affects context in two other ways: it adds the act of assertion to the context set (i.e., the set of shared presuppositions of the conversation) and reduces it by eliminating possibilities incompatible with what is asserted. The goal of the assertion is to get the audience to exclude possibilities incompatible with what you said. In terms of the two-dimensional formal account, the goal is to reduce the context set in a single, non-ambiguous, way.

Three pragmatic rules describe the way in which this goal is achieved: the proposition asserted must be true according to some, and false according to other, possibilities in the context set; a proposition must be expressed and have a truth-value relative to all possibilities in the context set; and the same proposition must be expressed relative to each possibility within the context set.<sup>17</sup> It might be that the original context defines a matrix with necessary propositions (i.e., violating



<sup>16</sup> Plantinga 1978: 134. Again, I have modified the original text to fit the examples.

<sup>17</sup> These “possibilities” are usually understood as “possible worlds” which represent possible ways for our world to be. They are not, in this sense, individual or partial possibilities of this or that thing being thus or so. They represent, rather, the totality of the world as being thus or so.

the first rule) and different ones relative to different worlds (i.e., violating the third rule). That is the case of matrix  $A$ . It gives us several candidates for the content of the assertion. We cannot pick a proposition because our direct rules of interpretation do not know yet tell us what was asserted.

If this happens, Stalnaker advises us to reinterpret the utterance, by taking the speaker to be saying something true. The diagonal proposition of the associated matrix represents the results of such reinterpretation. Formally speaking, this process is understood as “diagonalizing”. In the case of matrix  $A$ , the result of diagonalizing is matrix  $\dagger A$ .

$\dagger A$	W1	W2
W1	T	F
W2	T	F

Matrix  $\dagger A$  advises us to exclude possibility  $w2$  and keep  $w1$ , thereby reducing the context set in a single, unambiguous, way. Stalnaker (1978) argues that this model is good enough to explain the differences in informativeness. Consider our problematic sentences, say, (1a) and (1b).

- (1a) Mark Twain is Mark Twain.
- (1b) Mark Twain is Samuel Clemens.

I 87

Suppose I say (1b) to Andy, “Mark Twain is Samuel Clemens”. Now suppose there are two different possibilities as part of the context:  $w1$  where Mark Twain is Samuel Clemens and  $w2$  where ‘Mark Twain’ and ‘Samuel Clemens’ refer to two different people who happen to be close friends. This generates matrix  $B$  with two propositions, both of which are trivial.

$B$	W1	W2
W1	T	T
W2	F	F

How do I manage to say something informative? Stalnaker answers:

Now if we try to bring the initial context set into conformity with the third principle by shrinking it, say by throwing out world  $[w2]$ , we will

bring it into conflict with the first principle by making the assertion trivial. But if we look at what is actually going on in the example, if we ask what possible states of affairs the speaker would be trying to exclude from the context set if he made that statement, we can work backward to the proposition expressed. A moment's reflection shows that what the speaker is saying is that the actual world is [ $w1$ ] and not [ $w2$ ]. What he means to communicate is that the diagonal proposition of the [original] matrix [...] is true. (Stalnaker 1978: 91)

This explains how (1b) is informative and (1a) trivial and, thus, how these two sentences differ in informativeness. Assertions determine two-dimensional matrices that must be compatible with the speaker's presuppositions. Asserting (1b) generates a matrix with necessarily true and necessarily false propositions, thus giving place to a contingent diagonal. This diagonal proposition is informative. Asserting (1a) generates a matrix with only necessarily true propositions, thus giving place to a necessarily true diagonal. Neither this nor the horizontal propositions of the matrix are informative. Thus, the difference in informativeness. End of story.

Does this same story explain substitution failure? It seems it does. Remember the puzzle. Suppose I utter (5b) because I want to report Andy's beliefs to Victor.

88 |

(5b) Andy does not believe that Samuel Clemens wrote *Adventures of Huckleberry Finn*.

Suppose that  $w1$  is the actual world,  $w2$  is a world in which 'Mark Twain' and 'Samuel Clemens' refer to different people (none of which is the actual Mark Twain) and Mark Twain wrote *Adventures of Huckleberry Finn*, and  $w3$  is similar to  $w2$  but here Samuel Clemens wrote *Adventures of Huckleberry Finn*. This gives us matrix  $C$  for the embedded sentence 'Samuel Clemens wrote *Adventures of Huckleberry Finn*' in (5b).

$C$	$W1$	$W2$	$W3$
$W1$	T	F	F
$W2$	F	F	T
$W3$	F	F	T

Now consider the case of (5a), which determines matrix  $D$  for the embedded sentence 'Mark Twain wrote *Adventures of Huckleberry Finn*'.

(5a) Andy believes that Mark Twain wrote *Adventures of Huckleberry Finn*.

<i>D</i>	<i>W1</i>	<i>W2</i>	<i>W3</i>
<i>W1</i>	T	F	F
<i>W2</i>	F	T	F
<i>W3</i>	F	T	F

Given this distribution over the possibilities in the context, we are asked to reinterpret both assertions as saying something true. As a result, (5b) asks us to exclude possibility *w2* and keep possibilities *w1* and *w3*, as illustrated by the diagonal of matrix *C*, whereas (5a) asks us to exclude possibility *w3* and keep possibilities *w1* and *w2*, as illustrated by the diagonal of matrix *D*.

This accounts for substitution failure by explaining why we cannot substitute ‘Mark Twain’ for ‘Samuel Clemens’ even if they are coreferential in the actual world (possibility *w1*). Assertions determine two-dimensional matrixes that must be compatible with the speaker’s presuppositions. The embedded sentence in (5b) conveys the proposition illustrated by the diagonal of matrix *C*. The embedded sentence in (5a) conveys the one illustrated by the diagonal of matrix *D*. The matrices differ, that is why the names are not substitutable. End of story.

1 89

Stalnaker (1987) is hesitant to put things the way I did. To extend the same solution to the problem of belief ascription, says Stalnaker, we need a matrix compatible not only with the presuppositions of the speaker but, also, with the presuppositions of the subject of ascription. And here is where the difference is supposed to be relevant: “When the beliefs of the subject are very different from the presuppositions of the speaker, it is not always obvious how this is to be done” (Stalnaker 1987: 126).

This, however, is (if at all) a problem for the two-dimensional proposal of Stalnaker. It is certainly not the problem we are interested in. Suppose, for the sake of the argument, that we find a way to solve this problem –i.e., a way to define a matrix compatible with both the speaker’s and the subject’s presuppositions.<sup>18</sup> What would be the solution to our substitution failure problem?

Suppose that I utter (5a) and (5b), with the corresponding substitution failure among the embedded sentences. When uttered in the actual world, they

---

<sup>18</sup> Stalnaker does seem to have a rather obvious way to do this: “The propositional concept we construct is one not for the sentence as O’Leary would use or understand it, but for the sentence as the speaker and addressee would use and understand it if they were in the possible worlds relative to which the propositional concept is being defined”. See Stalnaker 1987: 127.

contradict each other. What does Stalnaker's theory say about them? Define a matrix compatible with Andy's presupposition that *Mark Twain is not the same person as Samuel Clemens*. Include counterfactual possibilities, like  $w_2$  and  $w_3$  above. The resulting matrices represent different propositions expressed in different worlds. Diagonalization yields different derived matrixes for the different embedded sentences. They do not contradict each other, they simply express different, logically consistent, propositions.

It is clear how this accounts for the failure of substitution. If the embedded sentences express different propositions, the names are not substitutable. It should also be clear that the same story applies to the differences in informativeness. I utter (1a) and (1b). If the reinterpretation of (1a) gives a necessary truth and that of (1b) a contingent one –as I have shown–, then (1b) is informative and (1a) is not.

So much for Stalnaker's account of differences in informativeness and substitution failure. If one problem is solved, the other is too. And vice versa as well. One more reason to think they go hand-in-hand.

#### 4. A historical account

90 |

I hope to have convinced you that *informativeness* and *substitution failure* are intimately related. Most philosophers in the tradition have failed to notice how strong and important this relation is. Surprisingly or not, it has been unnoticed by some and, at most, simply assumed by others.

Frege (1892), for example, seems to be a paradigmatic case of those who take informativeness and substitution failure to be two separate issues. As I have shown, his two solutions are completely different, even though each could have been generalized to account for the remaining case. The same seems to happen with Recanati (1993). He offers a solution to substitution failure and says close to nothing about informativeness. Given that he is an advocate of direct reference, it would be strange for him to do this if he had thought the issues were so closely related.<sup>19</sup> Soames' (2002) case is different. His account of informativeness is closely related to his account of substitution failure. But the way he presents them, one of them being a problem posed by propositional attitudes but not the other, suggests that he takes them to be two different problems. The same goes for Lycan's (2006) recent introduction to the philosophy of proper names.<sup>20</sup> He agrees that

---

<sup>19</sup> Similar considerations apply to Barwise and Perry 1983, Crimmins and Perry 1989, and Salmon 1986.

<sup>20</sup> See also Braun 2007.

Millianism faces four different challenges, informativeness and substitution failure being two of them. He does not say if the problems are related, and certainly does not seem to think they are *closely* related at all.

Things are clearer with Russell (1905). He knows Frege's informativeness puzzle well enough. Nonetheless, he does not consider it among the puzzles that any good theory should solve; all he deals with is substitution failure. I am *not* claiming, or implying, either that Russell did not know there was some such puzzle or that he did not have a way of explaining it. It may very well be that Russell does not pay attention to this puzzle, say, because he thinks it does not arise for idealized languages. If such is the case, then so be it. If this historical view of Russell is correct then Russell did think that they were two different unrelated problems. I do not defend that Russell fails to account for the puzzle, but that he fails to notice that these two puzzles are so intimately related with respect to *natural* languages that it is hard to see why they should be treated as distinct puzzles.

According to Russell, Frege's puzzle is simply about "why it is often worthwhile to assert identity" (Russell 1905: 214). This suggests that Russell assumed that they were two different problems. The same goes on with Quine (1953). His goal is to deal with the problem of substitution failure, which he considers to be owed to the presence of referential opacity. There is no reason to think that Quine is unaware of Frege's informativeness puzzle, yet he says nothing about it, even though identity statements appear all over the text.

1 91

With Kripke (1979) things are a bit unclear. He presents a new puzzle, this time about belief, involving substitution failure, but does not offer any solution to any of, what I take to be, closely related problems. He seems to be concerned pretty much only with substitution failure, though it could be argued that he is assuming that informativeness is closely related to it.

Finally, there are other philosophers who seem to think that the problems are closely related, somehow.<sup>21</sup> Plantinga (1978) does think that a proper solution should account for both, but he clearly takes them to be two different problems that cannot be solved by non-Fregeans. Stalnaker offers pretty much the same solution to both problems, but he explicitly says that these are two separate issues, given that substitution failure presents one further obstacle having to do with the ascription of attitudes.<sup>22</sup> Jacob (2014) presents both issues as problems of intentionality. He takes them to be related to each other but says nothing about their relation, and it is fair to say that he takes them to be different issues.

There is, truly speaking, no general consensus about how to understand

---

<sup>21</sup> See Plantinga 1978, Stalnaker 1978 and 1987, and Jacob 2014.

<sup>22</sup> See Stalnaker 1978 and 1987.

differences in informativeness and substitution failure. Things are a bit worse when it comes to how these issues relate to each other. The literature suggests that there is no consensus that takes informativeness and substitution failure to be *substantially* different, just as much as it suggests that there is no agreement that takes them to be as intimately related as I have claimed. I hope the arguments I have presented may contribute to such an agreement.

### 5. *Why substitution failure?*

I have argued that, given our current understanding of the *phenomena*, differences in informativeness and failures of substitution go together. And their explanations, too. I have not claimed, or argued, that this is necessarily so. Things could be otherwise, but they are not. And it is important to see why this is so, for understanding the way in which these *phenomena* actually relate is central for understanding and, hence, explaining them.

Consider a world in which there are differences in informativeness, but no failures of substitution. How would that world look like? This is a world in which (1a)–(1b), and (2a)–(2b), differ in informativeness.

92 |

- (1a) Mark Twain is Mark Twain.
- (1b) Mark Twain is Samuel Clemens.
- (2a) Mark Twain wrote *Adventures of Huckleberry Finn*.
- (2b) Samuel Clemens wrote *Adventures of Huckleberry Finn*.

What does this imply? Recall our formulation of *informativeness*:

**INF:** For any pair of sentences *S1* and *S2*, which are not necessary falsehoods and which vary only in the use of different coreferential names, a normal subject *A*, and a context *c*, *S1* and *S2* differ in informativeness in *c* iff *A* believes one and does not believe the other in *c*.

If the formulation is correct, the claim –which we are assuming– that (1a) differs in informativeness from (1b) presupposes that *as a matter of fact* Andy assents to one and not the other. I say “assents” to avoid being tendentious. In this way we leave open the question about Andy’s mental states and, hence, the question of substitution failure.

Now, suppose that, for some unexplained reason, among the inhabitants of world *wδ* there is a linguistic practice according to which all coreferential terms



are substitutable *salva veritate* within attitudinal contexts. It is important to note that this is *only* a change in linguistic practice and is meant *not* to include any cognitive or psychological difference between human beings here in the actual world and the inhabitants of  $w\delta$ . In particular, suppose that speakers in this world use the term ‘schmelieves’ that we define as:

Schmelieve: For any subject  $A$  and any names ‘ $m$ ’ and ‘ $n$ ’; if  $m=n$  then  $A$  schmelieves that  $m$  is  $F$  iff  $A$  schmelieves that  $n$  is  $F$ .

Suppose, furthermore, that we have really good evidence to the point that Andy schmelieves (1a). Hence, (4a) follows and, given that (4c) is true, (4d) follows too:

(4a) Andy schmelieves that Mark Twain is Mark Twain.

(4c) Mark Twain is Samuel Clemens.

(4d) Andy schmelieves that Mark Twain is Samuel Clemens.

Is there anything strange about this situation? A world like this is certainly metaphysically and logically possible. But is it cognitively possible? That is to say, can human subjects have mental representational states that match the uses of ‘schmelieves’ without thereby losing anything essential to human cognition? The answer, I believe, is negative. To illustrate it, let me begin by pointing at some surface oddities of ‘schmelieves’.

1 93

Ascribing schmeliefs to subjects may very well help illuminate their behavior. If all you know about *Adventures of Huckleberry Finn* is that Samuel Clemens wrote it, and you wonder why Andy has so many different editions of *Adventures of Huckleberry Finn* I can explain his behavior by saying: *Andy schmelieves that Samuel Clemens is the best author of the American tradition.* Yet, ascribing schmeliefs is, by far, not as explanatory as ascribing beliefs. It cannot explain why, for example, normal agents assent to one but not the other of the problematic pair of sentences (1a)-(1b). Andy’s distinctive assent to the distinct sentences is not an inference from the ascriber’s use of ‘schmelieves’; it is a puzzling fact that cannot be explained unless ‘schmelieves’ allows for some hyperintentionality.

To further illustrate this point, consider the following extension of the scenario above. Andy loves *Adventures of Huckleberry Finn* and he wants to meet the author. One day he is told that Samuel Clemens will be presenting a book in the coffee shop downstairs. He assents to both ‘Mark Twain wrote *Adventures of Huckleberry Finn*’ and ‘Mark Twain is Mark Twain’, but not to ‘Mark Twain is Samuel Clemens’. Given this, the definition of ‘schmelieves’ above, and the fact that Mark Twain=Samuel Clemens, we should conclude that Andy schmelieves that Mark Twain will be presenting a book in the coffee shop downstairs. Now, suppose further on, that Andy assents to ‘Samuel Clemens is the worst writer of

the American tradition'. Given this situation, it is terribly unclear what we can do (if anything at all) to understand and predict Andy's behavior on the basis of schmelief ascription. Should we expect him to go to the coffee shop downstairs? After all, Andy schmelieves that Mark Twain is the best writer of the American tradition and he schmelieves that he will be presenting a book downstairs. Yet, Andy was told that Samuel Clemens would be the one presenting and he does assent to 'Samuel Clemens is the worst writer of the American tradition'. Should we, then, expect him not to go downstairs? We have, I'm afraid, no clue of what to do.

To see what is missing, both linguistically and cognitively, we can begin by looking at ordinary belief-ascriptions. The natural result, in the scenario above, is to take Andy to believe that Samuel Clemens will be presenting a book downstairs, that Samuel Clemens is not Mark Twain, and that Samuel Clemens is the worst writer of the American tradition. A natural prediction would follow: Andy will not go downstairs to look for whoever is presenting the book. This illuminates both why Andy assents to one but not both sentences in (1a)-(1b) as well as his, otherwise strange, behavior with respect to Mark Twain and his book presentation. Linguistically speaking, we are missing a more illuminating attitude verb. Cognitively speaking, we would be missing more.

94 | If we take the adequacy of the belief report seriously, it suggests that human cognition –mental representational states in particular– can be guided by more than just the truth-conditional contents of the representations it makes use of. Schmeliefs are all about strictly following this truth-conditional content, yet, it seems clear that they cannot capture all there is to be captured by the mental representational states guiding human behavior. What else is there, then? What would be missing cognitively if we were to have just the cognitive correlate of non-hyperintentional attitude verbs such as 'schmelieve'?

Leslie (1987) has forcefully argued that one of our characteristically human cognitive abilities, that of pretending that such and such is the case, depends upon a more general ability to split representations of the surrounding environment from their referential, existential, and semantic conditions.<sup>23</sup> This cognitively general ability, available for any representational purpose, allows humans to form higher order representations (e.g., beliefs about beliefs and desires) and what he calls "decoupled" representations (e.g., the cognitive correlate of a quoted sentence with no assigned content), both of which can successfully guide an individual's behavior.

Consider, for example, the case of pretense. Children as young as three years

---

<sup>23</sup> See Leslie 1987, 1994 and 2002.

of age are able to engage in games of make-believe that make use of objects as props. Leslie (1987) considers cases of two and three-year-olds that can pretend of an empty cup that it is full. If the truth-conditional content of the representation ‘that cup is full’ were all that mattered for behavior-guiding purposes, it would simply be impossible for children to pretend in the mentioned way.

This ability is, of course, not unique to children, nor is it limited to pretending. Leverrier is famously known for having hypothesized both that *Vulcan perturbed the orbit of Mercury* and that *Neptune perturbed the orbit of Uranus*. He had similar kinds of evidence to accept the existence of both alleged planets, Vulcan and Neptune. Yet, as it turned out to be, there is no Vulcan. Leverrier presumably had a representational mental state of the form *Vulcan is a planet* as well as one of the form *Neptune is a planet*. Both mental states were, we may assume, of the same kind. The fact that ‘Vulcan is a planet’ is false (perhaps even necessarily false) did not preclude Leverrier from entertaining such representation and guiding his behavior accordingly.

Consider now what would be the case if all we were to have were schmelieve-like mental representational states. Suppose, for example, that Leverrier’s assent to “Vulcan is a planet” is good evidence of his schmelieving that *Vulcan is a planet*. Yet, as a matter of fact, ‘Vulcan’ has no referent, Vulcan is nothing whatsoever. It follows, then, that Leverrier’s schmelief is tantamount to schmelieving that *nothing is a planet* or, alternatively, to schmelieving nothing whatsoever. It seems that, on this view, Leverrier could not have had the representational mental state that, we assume, he did have. This result would extend to a good number of ordinary mental representational states –what would be a case of false belief on this view?– but also a good number of scientific hypothesis –e.g., phlogiston, ether, etc.

Leslie himself claims that linguistic “opacity” (what I’ve been calling hyperintensionality) has an important cognitive correlate: mental representational states that exhibit “object substitution” (see Leslie 1987: 416). Object substitution is most apparent in cases of basic pretense where the relevant mental state remains unaffected if the pretendedly referred-to object gets substituted by another one. Understanding theatrical representations is a good example. It doesn’t matter who gets to play the role of Hamlet in order to understand Shakespeare’s drama. If all there were to mental representational states were non-hyperintensional mental states such as schmelieve, there would simply be no pretense of this sort. For schmelieve, by definition, does not allow for object substitution. In general, fictional mental states like pretending, imagining, and perhaps hallucinating, don’t seem to be available any more.

I now want to go back to the initial question of this section. Can human subjects have mental representational states that behave the way schmeliefs do –i.e., can they do without any hyperintensional-like mental states– without

loosing anything central to human cognition? If Leslie is correct, then “mental hyperintensionality” turns out to be a design feature of human cognition.<sup>24</sup> Getting rid of such mental states by keeping only states of the kind of schmeliefs does seem to be like letting go a central part of human cognitive architecture. If so, and assuming attitudinal verbs are part of a broader machinery for explaining human cognition and behavior, it is not surprising that a good number of attitudinal verbs are hyperintensional. These, I hope, are good enough reasons to think we do need something else than just schmeliefs and schmelief-ascription. A more fleshed-out account of why we have hyperintensionality (linguistic and mental) is desirable, but beyond the scope of this paper. Suffice it to say that a world like  $w\delta$ , with informativeness but no substitution failure, seems to require some substantial departure from what seems to be actual human psychology.

### 6. A working hypothesis

96 | In the previous section I talked about “mental hyperintensionality” in reference to Leslie’s (1987) account of pretense. One of the defining features of this ability, according to Leslie, is the *phenomenon* of object substitution: certain mental states allow for the substitution of objects they allegedly refer to without thereby modifying the mental state. Leslie (1987) describes these mental states as decoupled from their referential, semantic, and existential conditions. He further exemplifies them by analogy with sentences. There is an obvious sense in which the sentence “Hamlet is the Prince of Denmark” remains the same regardless of whether ‘Hamlet’ is used to refer to David Kaplan or David Lewis. Now consider mental states that are concerned with the cognitive correlate of that English sentence. Like the sentence, the mental state will remain the same regardless of what the reference of ‘Hamlet’ is. Now, if the referent can be considered part of the semantic conditions of that mental state (or of that sentence), then that mental state can be said to be concerned with something else (whatever that may be) than its semantic content.

Now, consider what happens if, instead of object substitution, we have sign substitution. The very sense in which “Hamlet is the Prince of Denmark” remains the same regardless of what ‘Hamlet’ refers to is the sense in which *S1*: “Mark Twain wrote *Adventures of Huckleberry Finn*” and *S2*: “Samuel Clemens wrote *Adventures of Huckleberry Finn*” are not the same regardless of the reference of ‘Mark Twain’ and ‘Samuel Clemens’. In other words, *S1* and *S2* differ even if

---

<sup>24</sup> See Leslie 1987, 2002 and Friedman and Leslie 2007.

‘Mark Twain’ and ‘Samuel Clemens’ are coreferential. If we follow the analogy with mental states, the result is that cognitive correlate of *S1* will *not* remain the same if we substitute ‘Mark Twain’ for ‘Samuel Clemens’, even if the names are coreferential. And this will be so precisely because this mental state is concerned with something else (whatever that may be) than just its semantic content. Call these mental states “hyperintensional”.

The ability to have hyperintensional mental states, I believe, sheds light on our problematic *phenomena: informativeness* and *substitution failure*. Differences in informativeness are problematic when they appear among sentences that vary in the use of different coreferential terms. Hence, differences appear even where there are no semantic differences. If our mental states were to map only into semantic content, this would be a mystery. If, however, mental states can also be concerned with something else than just semantic content, this becomes a feature of human cognition. The sentences do not differ in their content, but they differ in the signs they use to convey that content. And it may very well be that this difference of signs is that “something else” that guides the hyperintensional state.

Recall our formulation of the problem

**INF:** For any pair of sentences *S1* and *S2*, which are not necessary falsehoods and which vary only in the use of different coreferential names, a normal subject *A*, and a context *c*, *S1* and *S2* differ in informativeness in *c* iff *A* believes one and does not believe the other in *c*.

197

The claim was: if a normal speaker assents to (or believes) one sentence without assenting to (or believing) the other, we have reasons to think there are differences in informativeness. The possibility of hyperintensional states, of mental states concerned with something else than just semantic content (i.e., the signs themselves), explains why these different attitudes can take place without there being semantic differences.

The same goes for substitution failure.

**SUBS:** For any two coreferential names ‘*N*’ and ‘*M*’, any context *c*, and a normal subject *A*: ‘*N*’ and ‘*M*’ are subject to substitution failure within reports in *c* of *A*’s first-order belief states iff *A* does not believe that ‘*N is M*’ is true in *c*.

The claim was: two coreferential names fail to substitute within attitudinal contexts whenever the relevant subject fails to accept the truth of the corresponding identity statement. Once more, the possibility of hyperintensional states explains why this can be the case. When the relevant mental states are concerned with the signs, then different signs, in this case different names, will

not be interchangeable. The speakers' ignorance that a certain identity statement is true is not a surprise: they seem to ignore that these representations are, for all they care, interchangeable.

This suggests that what gives place to both problematic *phenomena* is the presence of different names, full stop. It also suggests a new way to describe the problem that I would like to propose as a working hypothesis.

PUZZLE: For any two coreferential names 'N' and 'M', any context *c*, and any subject *A*: there will be a difference in informativeness between sentences *S1* and *S2*, for *A* in *c* iff 'N' and 'M' fail to be substitutable *salva veritate* within reports in *c* of *A*'s first-order belief states.

or alternatively

NAME PUZZLE: Any two *different* coreferential names 'N' and 'M', will: (i) give place to differences in informativeness; and (ii) fail to be substitutable within 'that'-clauses iff the associated attitudes of the relevant subject *A* are hyperintensional.

98 |

On this view, it is the presence of different names that gives place to an explanation of both informativeness and substitution failure *phenomena*. More specifically, what I want to propose is that a proper account of the *phenomena* must be based on a proper understanding of the cognitive processes involved in the use and acquisition of proper names, together with a more substantial understanding of what has come to be known as the theory of mind mechanism (see Leslie 1987).<sup>25</sup>

What these accounts suggests is, first, that the presence of different names opens the possibility of forming different mental representational states (see Hall 1996), and, second, that humans are able to manipulate mental representations in order to produce higher order representations (see Leslie 1987). With these two elements in hand, we can offer a sketch of an account of the *phenomena*. The presence of different names triggers the formation of different attitudes. The theory of mind mechanism (ToMM) gets the work done by manipulating representations and producing the relevant hyperintensional representations. All of this is meant to take place independently of the semantics (including reference) of the corresponding linguistic representations.

—

<sup>25</sup> See Valentine, Brennen and Brédart 1996, and Hall 1996.

With this sketch at hand, it is useful to compare the proposals with two other similar accounts owed to Fodor and Recanati, both of which claim that the relevant *phenomena* must be explained in virtue of differences among representations or signs. There are at least two levels at which there are important differences between my proposal and the available accounts.

First, an initial difference has to do with the fact that none of the available accounts seems to countenance the tight relation between informativeness and substitution failure. On the one hand, Fodor (1990) offers his account of substitution failure in terms of differences in belief states –without differences in content. Throughout, however, the problem of informativeness is ignored. Recanati (2012), on the other hand, does intend to offer an account of both *phenomena*, but he treats them as independent ones having to do with recognition and trading upon identity (Recanati 2012: 85) and with coreferentiality (Recanati 2012: 221). As Onofri (2014) shows, Recanati himself seems to have trouble for not considering these are closely related *phenomena*.

Second, there are important differences concerning the specific features of the accounts. Fodor (1990) claims that substitution failure cases are evidence not of differences in content but of belief states. On his view, “it must be that distinct belief-states can have the same content, i.e., there must be more to the identity of an attitude than its content and its mode”, and adds “differences in their vehicle’s seem to be all that’s left” (Fodor 1990: 168). Fodor adds little more to his claim that it is the vehicle of the belief that is to be blamed for the failure of substitution. Without any more specific details, it is difficult to judge whether and how this account is compatible with the one I sketched above. It seems, at first glance, to be similar to my claim that a difference of names is responsible for the differential belief formation. Yet, nothing is said concerning the processing and manipulation of mental representations, or anything that may explain why a normal subject may form the relevant hyperintensional states.

Recanati (2012) does offer a full-blown account, intended to be a referentialist friendly version of Frege’s theory by turning Frege’s senses into mental files that, nonetheless, do not determine semantic differences. On Recanati’s view, proper names and objects are related via a nondescriptive mode of presentation (Recanati 2012: 34). Nondescriptive modes of presentation are mental files that stand in epistemically rewarding relations with objects of reference. On this account, the presence of substitution failure is somehow related to the fact that subjects have different mental files associated with different names.

It is unclear how it is that Recanati’s account manages to account for both, informativeness and substitution failure, or if it does succeed (see Onofri 2014). But assuming that it does, the differences between this and the proposal I have presented should be clear. The view I have presented does not appeal to

mental files, epistemically rewarding relations and, most importantly, it does not appeal to *differences* among mental files. It is difficult to determine when two files become one, it is unclear whether a subject may have two files with exactly the same information in them and associated with exactly the same name, and thus it is unclear what amounts to a difference in the individuation of mental files that, in turn, gives place to substitution failure.

On the view I have proposed, there is no need to appeal to such differences. The proposed account is mainly based on how it is that normal subjects process proper names and, more importantly, on how they manipulate representations to form hyperintensional states.

This proposal prompts many questions, all of which I shall leave unanswered here. Which things count as different names? What counts as a hyperintensional mental state? What needs to be the case, cognitively speaking, for there to be differences in informativeness and substitution failure relative to a normal subject *A*? All of these should be addressed by a substantial account of the *phenomena* in question. I do not intend to do such thing in this paper. I'll be content to have shown that the apparently distinct problematic *phenomena* are intimately related, and why this is so.<sup>26</sup>

## REFERENCES

- Barwise, J. and Perry, J.** (1983), *Situations and Attitudes* (Cambridge: The MIT Press).
- Bealer, G.** (1993), "A Solution to Frege's Puzzle", *Philosophical Perspectives*, 7, *Language and Logic*: 17-60.
- Beaney, M.** (1997) (ed.), *The Frege Reader* (Oxford: Blackwell Publishing).
- Boolos, G.** (1990), *Meaning and Method: Essays in Honor of Hilary Putnam* (Cambridge University Press).
- Braun, D.** (2006), "Illogical but Rational", *Nous*, 40: 376-9.
- Braun, D.** (2007), "Names and Natural Kind Terms", in Lepore and Smith (2007: 490-515).
- Crimmins, M. and Perry, J.** (1989), "The Prince and the Phone Booth", *Journal of Philosophy*, 86: 685-711.

---

<sup>26</sup> The author would like to thank an anonymous referee for his/her helpful comments. Thanks are also owed to Axel Barceló and Lenny Clapp. The research for this paper was supported by PAPIIT-IA400112 and CONACYT CCB-2011-166502 research projects. The paper was completed as a visiting scholar at the UBA, Buenos Aires, Argentina. Thanks are owed to CONACYT (Of. 233188) and DGAPA (UNAM), PASPA program, for granting the needed scholarships.



**Devitt, M. and Hanley, R.** (2006) (eds.), *The Blackwell Guide to the Philosophy of Language* (Oxford: Blackwell).

**Fodor, J.** (1990), "Substitution Arguments and the Individuation of Beliefs", in Boolos (1990:63-79).

**Frege, G.** (1879), "Begriffsschrift", in Beaney (1997: 47-78).

**Frege, G.** (1892), "On Sense and Nominatum", in Martinich (1996: 186-198).

**Frege, G.** (1917), "The Thought", in Beaney (1997: 325-345).

**Friedman, O. and Leslie, A.M.** (2007), "The Conceptual Underpinnings of Pretense: Pretending is not 'Behaving as If'", *Cognition*, 105: 103-24.

**Hall, D.G.** (1996), "Preschooler's Default Assumptions about Word Meaning: Proper Names Designate Unique Individuals", *Developmental Psychology*, 32: 177-186.

**Jacob, P.** (2014), "Intentionality", *The Stanford Encyclopedia of Philosophy*, (Winter 2014 Edition), Edward N. Zalta, (ed.), URL = <<http://plato.stanford.edu/archives/win2014/entries/intentionality/>>

**Kripke, S.** (1979), "A Puzzle about Belief", in Martinich (1996: 382-409).

**Kripke, S.** (1980), *Naming and Necessity* (Cambridge: Harvard University Press).

**Kroon, F.** (2004), "Descriptivism, Pretense, and the Frege-Russell Problems", *Philosophical Review*, 113: 1-38.

**Lepore, E. and Smith, B.** (2007) (eds.), *Handbook of Philosophy of Language* (Oxford: Oxford University Press).

**Leslie, A.M.** (1987), "Pretense and Representation: The Origins of "Theory of Mind"", *Psychological Review*, 94: 412-426.

**Leslie, A.M.** (1994), "Pretending and Believing: Issues in the Theory of ToMM", *Cognition*, 50: 211-238.

**Leslie, A.M.** (2002), "Pretense and Representation Revisited", in Stein, Bauer and Rabinowitz (2002: 103-114).

**Lewis, D.** (1981), *On the Plurality of Worlds* (Oxford: Blackwell Publishing).

**Lillard, A.S.** (2002), "Just through the Looking Glass: Children's Understanding of Pretense", in Mitchell (2002: 102-114).

**Lycan, W.** (2006), "Names", in Devitt and Hanley (2006: 255-273),

**Marcus, Ruth B.** (1981), "A Proposed Solution to a Puzzle about Belief", *Midwest Studies in Philosophy*, 6: 501-510.

**Martinich, A.P.** (1996) (ed.), *The Philosophy of Language* (Oxford: Oxford University Press).

**McKay T. and Nelson, M.** (2010), "Propositional Attitude Reports", *The Stanford Encyclopedia of Philosophy*, (Spring 2014 Edition), E.N. Zalta, (ed.), URL = <<http://plato.stanford.edu/archives/spr2014/entries/prop-attitude-reports/>>.

**Mitchell, R.W.** (2002) (ed.), *Pretending and Imagination in Animals and Children*, (Cambridge: Cambridge University Press).

**Nichols, S. and Stich, S.** (2000), "A Cognitive Theory of Pretense", *Cognition*, 74, 115-47.

**Nichols, S. and Stich, S.** (2003), *Mindreading: An Integrated Account of Pretense, Self-Awareness and Understanding Other Minds* (Oxford: Oxford University Press).

**Onofri, A.** (2014), "Mental Files and Rational Inferences", *Inquiry*, 4: 1-15.

**Plantinga, A.** (1978), "The Boethian Compromise", in Plantinga (2003: 122-138).

- Plantinga, A.** (2003), *Essays in the Metaphysics of Modality* (Oxford: Oxford University Press).
- Quine, W.V.O.** (1953), *From a Logical Point of View* (Cambridge: Harvard University Press).
- Quine, W.V.O.** (1953), "Reference and Modality", in Quine (1953: 139-159).
- Recanati, F.** (1993), *Direct Reference* (Oxford: Blackwell).
- Recanati, F.** (2012), *Mental Files* (Oxford: Oxford University Press).
- Russell, B.** (1905), "On Denoting", in Martinich (1996: 199-207).
- Salmon, N.** (1986), *Frege's Puzzle* (Cambridge: The MIT Press).
- Salmon, N.** (1989), "Illogical Belief", *Philosophical Perspectives*, 3: 243-285.
- Salmon, N.** (2006), "The Resilience of Illogical Belief", *Nous*, 40: 369-375.
- Schiffer, S.** (1987), "The 'Fido'-Fido Theory of Belief", *Philosophical Perspectives*, 1, 455-80.
- Schiffer, S.** (2006), "A Problem for Direct Reference Theory of Belief Reports", *Nous*, 40: 361-368.
- Soames, S.** (1987), "Direct Reference, Propositional Attitudes, and Semantic Content", *Philosophical Issues*, 15: 47-87.
- Soames, S.** (2002), *Beyond Rigidity* (Oxford: Oxford University Press).
- Stalnaker, R.** (1978), "Assertion" in Stalnaker (1999: 78-95).
- Stalnaker, R.** (1987), "Semantics for Belief" in Stalnaker (1999: 117-129).
- Stalnaker, R.** (1999) *Context and Content* (Oxford: Oxford University Press).
- Stein, N.L., Bauer, P.J. and Rabinowitz, M.** (2002) (eds.), *Representation, Memory and Development: Essays in Honor of Jean Mandler* (Mahwah: Lawrence Erlbaum).
- Swanson, E.** (2006), "Frege's Puzzle and Pragmatic Presupposition" in Swanson (2006: 9-36).
- Swanson, E.** (2006) *Interactions with Context* (PhD Dissertation) (Cambridge: MIT).
- Valentine, T., Brennen, T. and Brédart, S.** (1996), *The Cognitive Psychology of Proper Names* (London: Routledge).
- Wettstein, H.** (1989), "Turning the Tables on Frege or How is it that "Hesperus is Hesperus" is Trivial?", *Philosophical Perspectives*, 3: 317-339.

*Recibido: 05-2014; aceptado: 12-2014*