UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS

MATHEUS VALENTE LEITE

Singular Content and *De Jure* Coreference

Conteúdo Singular e Correferência *De Jure*

CAMPINAS
2016

MATHEUS VALENTE LEITE

SINGULAR CONTENT AND DE JURE COREFERENCE

CONTEÚDO SINGULAR E CORREFERÊNCIA DE JURE

Dissertação apresentada ao Instituto de Filosofia e Ciências Humanas da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Filosofia.

Dissertation presented to the Institute of Philosophy and Human Sciences of the University of Campinas in partial fulfillment of the requirements for the degree of Master, in the area of Philosophy.

Supervisor/Orientador: Prof. Dr. Marco Antonio Caron Ruffino

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELO ALUNO MATHEUS VALENTE LEITE, E ORIENTADO PELO PROF. DR. MARCO ANTONIO CARON RUFFINO.

CAMPINAS

2016

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Filosofia e Ciências Humanas
Cecília Maria Jorge Nicolau - CRB 8/3387

V234s	Valente, Matheus, 1990-
	    Singular content and de jure coreference / Matheus Valente Leite. –
	Campinas, SP : [s.n.], 2016.
	
	    Orientador: Marco Antonio Caron Ruffino.
	    Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de
	Filosofia e Ciências Humanas.
	
	    1. Frege, Gottlob, 1848-1925. 2. Recanati, François, 1952-. 3. Referência
	(Filosofia). 4. Linguagem - Filosofia. I. Ruffino, Marco,1963-. II. Universidade
	Estadual de Campinas. Instituto de Filosofia e Ciências Humanas. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Conteúdo singular e correferência de jure
Palavras-chave em inglês:
Reference (Philosophy)
Language - Philosophy
Área de concentração: Filosofia
Titulação: Mestre em Filosofia
Banca examinadora:
Marco Antonio Caron Ruffino
Ludovic Soutif
Eros Stefano Corazza
Data de defesa: 04-07-2016
Programa de Pós-Graduação: Filosofia

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS

A Comissão julgadora dos trabalhos de Defesa e Dissertação de Mestrado, composta pelos Professores Doutores a seguir descritos, em sessão pública realizada em 04/07/2016, considerou o candidato Matheus Valente Leite aprovado.

The Committee judging the Defense and Master's Dissertation, composed by the following Professors, in a public session realized at the 04/07/2016, has approved the candidate Matheus Valente Leite.

Prof. Dr. Marco Antonio Caron Ruffino

Prof. Dr. Eros Stefano Corazza

Prof. Dr. Ludovic Soutif

A Ata de Defesa, assinada pelos membros da Comissão Examinadora, consta no processo de vida acadêmica do aluno.

Dedico a Rita, minha mãe.

ACKNOWLEDGEMENTS

FUNDING ACKNOWLEDGEMENT

FINANCIAMENTO

RESUMO

Esta dissertação é composta de três capítulos que, apesar de tratarem sobre temas distintos, igualmente pressupõem que o conceito de coreferência *de jure* deve estar no cerne das discussões sobre conteúdo e representação singular. No capítulo 2, tentarei reabilitar o principal argumento de Frege (1948) em favor da existência de Sentidos em face da acusação de circularidade recentemente posta por Glezakos (2009). Argumentarei que as diferenças de valor cognitivo entre sentenças, como "Cícero é Cícero" e "Cícero é Túlio", se reduzem à ocorrência (ou não ocorrência) da relação de coreferência *de jure* entre seus tokens referenciais. Já que, como mostrarei, o reconhecimento desta relação não pressupõe qualquer teoria semântica substancial, ele pode ser usado para fundamentar a introdução de Sentidos. O capítulo 3 avalia caso a introdução de Sentidos tem sucesso em explicar a coreferência *de jure* já que a última foi demonstrada ser não-transitiva (Fine 2007, Pinillos 2011), ao contrário da identidade de Sentidos. Argumentarei que a postulação de Sentidos pode fundamentar uma relação não-transitiva, já que Sentidos podem instanciar relações mais refinadas que a mera identidade ou diferença, e.g. possuir uma parte comum. Finalmente, o último capítulo contém uma discussão sobre conteúdo singular em cenários diacrônicos envolvendo expressões indexicais. Defenderei a tese de que teorias que rejeitam a factividade, i.e. a tese de que identidade de conteúdo pressupõe coreferência, não são superiores a teorias que rejeitam a transparência, i.e. a tese de que a identidade de conteúdo é conhecível *a priori*. Já que temos explícitas razões para manter estas duas teses, teremos que adotar uma teoria como a de Recanati (2012, 2015, forthcoming), de acordo com quem a relação de tokens indexicais diacrônicos é geralmente mais fraca que a identidade de conteúdo, porém mais forte que a mera diferença. Todavia, argumentarei que Recanati erra em alguns detalhes, especialmente sobre se pode haver identidade de conteúdo entre representações memoriais e aquelas originando estas memórias. Meu ponto final é que, se meus argumentos são corretos, deveremos distinguir a identidade de modos de apresentação da identidade de valor cognitivo, a primeira sendo mais refinada que a segunda.

Palavras-chave: Coreferência de jure, Sentidos, valor cognitivo, pensamento singular, dinâmica cognitiva.

ABSTRACT

This dissertation is composed by three chapters which, albeit dealing with distinct particular issues in the philosophy of language and mind, share the underlying assumption that the concept of *de jure* coreference must be at the forefront of theorizing about singular thought and representation. In chapter 2, I will try to rehabilitate Frege's (1948) master argument for the existence of Senses in face of the accusation of circularity recently posed by Glezakos (2009). I will argue that the differences of cognitive value between sentences, such as "Cicero is Cicero" and "Cicero is Tully", boil down to the occurrence (or lack thereby) of the relation of *de jure* coreference between their referential tokens. Since, as I will show, the acknowledgement of that relation does not presuppose any substantial semantic theory, it may logically be used to ground the introduction of Senses. Chapter 3 assesses whether the introduction of Senses is successful in explaining *de jure* coreference in view of the fact that the latter was claimed to be a non-transitive relation (Fine 2007, Pinillos 2011), as opposed to the relation of sameness of Sense. We will argue that the postulation of Senses can ground a relation that is non-transitive, since Senses can instantiate relations more fine-grained than mere identity or difference, e.g. having a part in common. Finally, the last chapter contains a discussion about singular content in diachronic scenarios involving indexical expressions. I will claim that theories rescinding from factivity, i.e. the claim that identity of content entails coreference, stand in no better footing than theories that rescind from transparency, i.e. the claim that identity of content is knowable *a priori*. Since we have explicit reasons to maintain these two claims, we might need to adopt a theory such as Recanati's (2012, 2015, forthcoming), according to which the relation between diachronic indexical tokens is usually weaker than identity of content, but stronger than mere distinctness of content. However, I will claim that Recanati is wrong about some important details, more particularly, about whether there can ever be proper identity of content between memory representations and the representations which they are a memory of. My final point is that, if my arguments are sound, we may need to distinguish the relation of identity of modes of presentation from cognitive value, the former being more fine-grained than the latter.

Keywords: Coreference de jure, Senses, cognitive value, singular thought, cognitive dynamics.

TABLE OF CONTENTS

1. INTRODUCTION

While most of the representations that we produce – whether linguistic or mental – seem to be accountable for by their truth conditions, singular representations exhibit semantic aspects that refuse to be reduced to their referents. This dissertation is about the *hyper-intensionality of singular content*, i.e. the fact that singular representations have semantic aspects that outstrip the referent of its constituents. Let me try to flesh that out in more details.

By "content", I mean the fundamental semantic entity which is associated to complex and simple representational devices, e.g. words, utterances, mental states etc. Contents are usually seem as corresponding to (i) what is expressed by an utterance of a sentence (ii) and what constitutes a representational mental state. It is a reification of the intentional aspect of a linguistic and/or mental representation. The content of an utterance is usually defined as a Proposition, although I will refrain from using that concept in favor of the Fregean-tinged concept of a Thought. The reason for this being merely to avoid the association that Propositions usually have to Russellian Propositions, i.e. contents individuated at the level of reference, or functions from possible worlds to extensions.

As it will be clear later on in this dissertation, I assume a pluralist view about the semantic contents of linguistic and mental representations. The claim that the pre-theoretical locution "what is expressed by an utterance" must correspond to exactly one semantical concept in every possible case is unwarranted and does not stand by itself. Similarly, our pre-theoretic practice of referring to *what someone is thinking* or *what someone believes* may point to multiple semantic aspects of mental representations, so that it is not surprising that multiple notions of semantic contents may be needed in order to explain it.

By "singular content", I am pointing to the fact that we will be concerned with the content of a restricted class of representations: those that are essentially directed towards a specific object. In natural languages, such as English or Portuguese, there is a class of expressions that we can use to achieve singularity: the referential expressions. The most common examples of referential expressions in natural languages are (i) proper names, such as "Aristotle" or "Julius Caesar", (ii) indexical expressions, such as "Today" or "Here", and (iii) demonstratives, such as "This", "That" or "He"[1]. The defining characteristic of those expressions is not that they have singular referents, but that they are *essentially* connected to them. I emphasize "essentially" to make it clear that it is their modal behavior that sets them apart from a descriptive term, such as a definite description. That is, even though both

---

[1] I'm leaving out the category of descriptive names although they will be useful for our later discussions.

"the most prolific student of Plato" and "Aristotle" could be said to refer to Aristotle[2], only the second still would if he had never met Plato in the first place.

Finally, by "hyper-intensionality", I mean nothing more than that the semantic aspects of a singular representation outstrips its intension. A brief digression on semantic concepts will help to expand on this point.

In formal semantics, the concept of an expression's *extension* corresponds to the simplest aspect of its truth-conditional content: it is the minimal contribution it offers to the truth-value computation of the sentence in which it is embedded. For referential terms in formal languages, such as individual constants, their extension is simply the object associated to them by means of the relevant assignment function. Similarly, for referential terms in natural languages, their extension is simply the object they actually refer to. Analogously, the extension of a predicate is a set of objects and the extension of a sentence is its truth-value. This all is founded on the Tarskian principle that a sentence of the logical form *a is F* is true if and only if the referent of "*a*" pertains to the set of F-things. Whenever a referential term appears in indicative contexts such as __ i*s F*, it seems that its truth-conditional implications are exhausted by its referent.

The problem with extensions is that they are too coarse-grained to individuate truth-conditional content in certain specific contexts. This limitation is very easy to spot in relation to modal contexts, where a certain content is evaluated in regards to its necessity or contingency. Since, for example, every chordate (creature with a heart) is, in fact, a renate (creature with a kidney), then the predicates "chordates" and "renates" are co-extensional. Nonetheless, they are not interchangeable in modal contexts such as "Necessarily, every _____ is a creature with a heart".

Extensions are modally naïve. Two representations that contingently refer to the same object have the same extension just as much as two representations that necessarily do. These considerations point to the need to consider the extension of expressions over the whole space of metaphysical possibilities, not only at the actual world.

An *intension*, in the modern sense that goes back to Carnap (1947), is then defined to be a function from possible worlds to extensions. While the most prolific student of Plato was Aristotle in the actual world, it could have been someone else. That is, there are possible worlds in which the description "the most prolific student of Plato" picks out someone different from Aristotle.

The proper name "Aristotle", on the other hand, since it is essentially associated to

---

[2]It is controversial whether definite descriptions should really be said to have singular referents or not. At the very least, if they are to be included in the class of referential expressions, the reference relation will have to be relativized to a particular possible world (see Evans 1979).

Aristotle, outputs the same referent in every possibility it confronts. This can be easily seen from the fact that the truth-conditions of a modal sentence involving "Aristotle", such as "Necessarily, Aristotle is the tutor of Alexander The Great", seem to supervene on whether Aristotle himself (as opposed to the satisfier of some description) was the tutor of Alexander The Great in every possible world.

In modal terms, identity of extension is identity of reference in the actual world, whereas identity of intension is identity of extension in every possible world.

Very unfortunately, not even intensions are fine-grained enough to individuate singular content, and this is precisely why I claimed these contents were said to be *hyper-intensional*. That fact is easily observed when one takes into account doxastic and/or epistemic contexts; for example, those provided in natural language by operators such as "believes that", "thinks that" or "knows that". The Roman orator Marcus Tullius Cicero, for example, is sometimes referred to by his name "Cicero", sometimes by his name "Tullius". As we have previously seen, proper names are essentially associated to their bearers, so that both "Cicero" and "Tullius" refer to Marcus Tullius Cicero in every possible world. Thus, if we construe their intensions as functions from possibilities to referents, they come out identical. However, if the singular content of these names were to be exhausted by their intensions, how could we make sense of the fact that sentences such as "Suzanne believes that Cicero is a Roman orator" and "Suzanne believes that Tullius is not a Roman orator" can have distinct truth-values? In other words, if the truth-conditional content of these proper names is exhausted by their intensions, and if their intensions are identical, it should not be possible that two sentences which differ merely in the substitution of one name for the other[3] would have distinct truth-values. However, the two aforementioned belief-ascriptions seem to contradict this claim.

At this point philosophers face the difficult decision between (i) admitting that singular representations have semantic aspects beyond their intensions or (ii) overriding our commonsensical intuitions about belief reports and claiming that both sentences about Suzanne's beliefs necessarily have the same truth-value.

Generally speaking, those who have chosen the first route have been closer to the Fregean tradition of singular content, while the others are called "referentialists", since they strived to keep their semantics at the level of intensional content. In this dissertation, I intend to take the former route and, thus, to align myself with what I take to be a broadly Fregean theory of singular content.

However, my understanding of "Fregeanism" shall be minimal. To be sure, I intend to present a constructive approach to Fregeanism according to which its properties will be individually

---

[3] I will simply disregard the possibility that the names embedded in the belief ascriptions are type-distinct from the two names of Marcus Tullius Cicero.

motivated and defended against an ever-increasing set of basic *data*. The starting point for our discussion will be the basic (although objectionable) *datum* that there can be truth-conditional differences between coreferential representations. Every investigation must start somewhere, and this *datum* seems as fair a choice as one could be.

One further preliminary remark. In this dissertation, I intend to move freely between linguistic- and mental-talk. The notion of semantic content, it should by now be clear, applies equally to linguistic representation as it does to mental representation and I intend to make no commitment in respect to one of those applications being more fundamental than the other. I mean by this no more than that when one utters the sentence "snow is white", one expresses a certain content (which may be identical or distinct from its truth-conditions) that can usually be reached by simple disquotation, e.g. the content that snow is white. That content, furthermore, is identical to the content one endorses when one believes that snow is white. In the linguistic case, an utterance is the vehicle for a certain content; analogously, in the mental case, a mental state, e.g. a belief, does the job of an utterance as far as being the vehicle of a determinate content. Whereas natural language contains a set of referential terms, mental representation contains a set of referential concepts. I will suppose that, for all relevant semantic purposes, each referential term has a mental analogue: (i) proper names stand in a one-to-one relation to name-like concepts, (ii) indexical and demonstrative terms stand in a one-to-one relation to indexical and demonstrative concepts etc[4]. More generally, when one thinks [at t^1] that R is F, for a certain referential concept R and property F, one endorses the same content one would have expressed had one uttered "R is F" at t^1.

A question arises about the disanalogies between linguistic entities and mental states: the former are concrete particulars and have logical structure/syntax while the latter possess an elusive ontology and do not seem to be structured in any obvious way. These questions are interesting and ground foundational questions in the Philosophy of Language and Mind but will be completely bypassed in this dissertation.[5]

As will be seen in future sections, sometimes it is easier to make certain remarks by saying that *one believes that p*, while, in other cases, it is better to take the longer route and say that *one expresses a certain content by uttering "p"*. In the end, my motivation for moving freely between linguistic- and mental-talk is mainly pedagogical: I think that the hyper-intensionality of singular

---

[4]In chapter 5, I will provide an argument in favor of the thesis that indexical expressions have conceptual analogues.

[5]If one is really uncomfortable with the idea that mental representation is analogous to linguistic representation, one could provisionally adopt the Language of Thought hypothesis (Fodor, 1975) and suppose that mental states are analyzable into syntactically structured Mentalese sentences in the Language of Thought, although I definitely don't intend that anything substantial depend on that. There are further questions concerning indexical representation but these will be postponed until later in this work.

representations is as much a puzzle in the Philosophy of Language as it is in the Philosophy of Mind even though many philosophers and students fail to appreciate this fact.

In chapter 2, I introduce the main role the concept of singular content has to play: the cognitive value role (2.1). I then describe how these issues were originally presented by Frege and advance a challenge, taking inspiration from Glezakos (2009), according to which Frege's narrative suffers from either incoherence or circularity. I claim that the Glezakos' understanding of Frege's Puzzles is too narrow (2.2). In the following section, I introduce the concept of *de jure* coreference and argue that we should understand Frege's talk of cognitive value in its terms (2.3); I then argue that this interpretation of Frege's argument can overcome Glezakos' challenge (2.4). In section 2.5, I present Campbell's (1987) regress argument with the intent of showing that *de jure* coreference does not presuppose any substantial theory of content, in particularly not Fregeanism. Finally, I draw some conclusions from the previous discussions and point the way to further issues (2.6).

In chapter 3, I discuss two general types of solution one could give to Frege's puzzle. Adopting Fine's (2009) terminology, I will call one type of solution "intrinsicalist" and the other "relationist". We will go over several reasons one would choose one type of solution in detriment of the other, focusing on a couple of arguments (Pinillos 2011) purporting to establish that *sameness of cognitive value* is a non-transitive relation, as opposed to the Fregean *sameness of Sense* relation. In section 3.1, we will analyze the "anaphorical case", and I will claim that there is a natural Fregean solution to the problem it poses, since Senses can instantiate logical relations more diverse than mere identity or difference, such as the relation of having a part in common. This will lead us to acknowledge a distinct type of coreferential relation that can be instantiated by referential tokens: *weak de jure* coreference. That relation is weaker than *de jure* coreference, but stronger than mere independence. I will argue that a relationist is in more trouble acknowledging that relation than the Fregean, so its existence might weigh in favor of the latter. Section 3.2 contains a discussion Frege's thesis that expressions shift their referents in opaque contexts and about *de dicto* and *de re* attitudes. In that section, I claim that there is no failure of transitivity of *de jure* coreference in the attitudinal representations offered by Pinillos. This chapter ends with a brief appendix showing how, once we acknowledge the existence of *de jure* coreference, four distinct relations of semantic identity can be defined for identity representation.

Chapter 4, the last one of the bunch, contains a discussion about singular content in diachronic scenarios involving indexical expressions. I will claim that theories rescinding from factivity, i.e. the claim that identity of content entails coreference, stand in no better footing than theories that rescind from transparency, i.e. the claim that identity of content is knowable *a priori*. Since we have explicit

reasons to maintain these two claims, we might need to adopt a theory such as Recanati's (2012, 2015, forthcoming), according to which the relation between diachronic indexical tokens is usually that of *weak de jure* coreference, i.e. weaker than identity of content, but stronger than mere distinctness of content. However, I will claim that Recanati is wrong about some important details, more particularly, about whether there can ever be proper identity of content between memory representations and the representations which they are a memory of. My final point is that, if my arguments are sound, we may need to distinguish the relation of identity of modes of presentation from cognitive value, the former being more fine-grained than the latter.

TERMINOLOGICAL REMARKS

I will use the expression "representations" to encompass tokens of both sentences and mental states (e.g. beliefs and thoughts) and I will use the expression "tokens" to abbreviate tokens of referential expressions or singular concepts.

Thus, the sentence "that representation contains two coreferential tokens" is to be read as an abbreviation of "that sentence/mental state token contains two coreferential tokens of referential expressions/singular concepts". My focus on tokens of sentences/mental states, as opposed to types, is by now a familiar move in the philosophical literature. Its motivation includes, but is not exhausted by, (i) the fact that the identity conditions of name's types in natural languages are not clear-cut (see next chapter for more on that) and (ii) the fact that indexical types cannot be assigned a content in the absence of a context of use.

This terminology will enable us to make points about singular content that are neutral between linguistic and mental representation.

2. FREGE'S PUZZLE[6]

In the introduction, we briefly went over the reasons for claiming that singular content is hyper-intensional, let's now take a more detailed look into those matters.

The point about hyper-intensionality was, in summary, that singular representations seem to have semantic aspects, e.g. those related to doxastic and epistemic contexts, which seem to be more fine-grained than their intensional content. It does not matter, for our purposes, whether one decides to characterize intensional content by means of Russellian Propositions or by means of sets of possible worlds[7]: both characterizations fall short of carving semantic space in correspondence with its cognitive implications. This is why Suzanne, for example, can be perfectly rational in believing that Cicero was a Roman orator without having any beliefs about Tully, i.e. because there are semantic differences between cointensional representations. Let us put that in the form of an axiom:

(A1) Two cointensional representations may differ semantically.

Two remarks about (A1). Firstly, it is a thesis that I am assuming instead of arguing for. Regardless of that, it is not a blind assumption and I believe the burden of proof belongs to its detractors. Notoriously, some philosophers, those sometimes called "referentialists", have denied the existence of any semantic aspects that are more fine-grained than intensional content. These philosophers have usually advanced theories locating our intuitions behind (A1) in the pragmatics of natural language (Salmon 1986, Soames 2002), the deep-level syntax of natural language (Fiengo & May, 2006) or the syntax of the Language of Thought (Fodor 2008, Heck 2012).

To be sure, very few philosophers seem to argue in favor (or in detriment) of (A1) by means of a transcendental argument. At the fundamental level, the adoption (or rejection) of (A1) seems to depend on how one fleshes out our commonsensical intuitions about attitude ascriptions. That is, the defenders of (A1) have usually pointed out that differences in epistemic/doxastic role seem to affect the truth-value of attitude ascriptions, and that, for this reason, they should supervene on a semantic basis. On the other hand, the detractors of (A1) have usually denied that attitude ascriptions are truth-conditionally sensitive to anything other than intensional content and explained our 'dubious' intuitions in some other way, e.g. our intuitions about attitude ascriptions' truth-values track pragmatic enrichments (Soames 2002), attitude ascriptions are sensitive to deep-level syntactic structure (Fodor

---

[6]The material in this chapter is reminiscent of my paper "Frege's Puzzle in Natural Language" co-authored with Emiliano Boccardi.
[7]Representing a proposition as a set of possible worlds (in which it is true) is equivalent to representing it as a function from possible worlds to truth-values.

2008) etc.

Since theorizing has to start at a certain point, I will simply assume that singular representations really have a hyper-intensional aspect which is shown by their behavior inside doxastic and/or epistemic contexts. Furthermore, the burden of proof seems to be on those that deny (A1) since their overriding our commonsensical intuitions about attitude ascriptions' truth-values cannot help but to sound revisionary in a very radical sense. Thus, I take it that if I manage to construct a good theory based on (A1), I will have automatically provided an argument in its favor.

Secondly, (A1) does not entail that intensional content is a useless theoretical notion. In fact, most linguistic and mental representation seems to be semantically exhausted by their intension. That is, if we were only concerned with extensional contexts, i.e. those where a singular representation's truth-value is determined by its extension (e.g. those created by the operator "It is actually true that"), or even modal contexts, i.e. those where contents are judged according to their necessity or contingency, we would never need to bother with anything more refined than extensions or functions from possible worlds to extensions. It is only when we are concerned with doxastic and/or epistemic representations that the need for a hyper-intensional notion of content makes itself evident[8].

The semantical picture that I am going to assume is pluralistic: representations can (and, usually, should) be assigned more than one notion of semantic content and these contents may have distinct theoretical uses. Thus, even if *intensional content* is useful for many purposes, we need at least one more distinct notion of semantic content in order to deal with doxastic and/or epistemic contexts.

While the *intensional content* of (the most natural interpretation of) "Cicero is Cicero" is identical to that of "Cicero is Tullius", they seem to be doxastically distinct. By that, I mean that one could rationally believe the former while disbelieving the latter. They also seem to be epistemically distinct, and by that, I mean that the first sentence seems to be knowable *a priori* while the other requires empirical evidence in its support.

As it should be becoming clear by now, these hyper-intensional aspects of singular representations seem to be intrinsically associated with their impact in a thinker's cognition and rationality, that is, with a representation's cognitive value. From now on, I will refrain from using the (admittedly ugly) concept of *hyper-intensional semantic aspects* and substitute it for the more familiar-sounding concept of the cognitive value of a representation.

The notion of cognitive value has its roots in Feigl's influential translation of Frege's *Über*

---

[8]In this dissertation, we will be concerned only with that fraction of hyper-intensionality that is related to doxastic and/or epistemic contexts. Naturally, the concept of hyper-intensionality is broader than that and extends to every case of logically equivalent representations that seem to differ in some semantic aspect, e.g. the truths of Logic, Mathematics, Philosophy (?) etc.

Sinn und Bedeutung where it is said that "'a = a' and 'a = b' are obviously statements of differing cognitive value" (1949, p. 209) because the first can be known *a priori* and is trivial while the second expresses valuable knowledge about the world. The general idea behind Frege's introduction of this concept is identical to the one that led us to claim that cointensional representations may differ in some semantic aspects: intensional content is too naïve too individuate cognitive value.

So much for those terminological refinements. Let us try to abstract something substantial from these points.

2.1 TWO TESTS FOR COGNITIVE VALUE

The aforementioned observations about "Cicero" and "Tully" seem to suggest two ways to compare representations concerning their cognitive value. They can be formulated as two tests: the doxastic and the epistemic tests.

According to the doxastic test[9], if it is not irrational to believe one representation *p* while disbelieving (or suspending belief from) *p'*, then *p* has a distinct cognitive value than *p'*. According to the epistemic test, if one representation is *apriori* while the other is *aposteriori*, then they have distinct cognitive values. The first test postulates a semantic difference between representations that one can rationally bear conflicting attitudes towards; the second test entails that no representations can be semantically identical if they are epistemically distinct (e.g. one is *apriori*, the other, *aposteriori*).

These two tests seem to point at the same phenomenon by means of different routes, but they are not equivalent. In fact, I will argue that there are two reasons to favor the doxastic test over the epistemic one. The first reason is that the doxastic test seems to be more informative than the epistemic one, since there are a handful of representations about which only the former test has something interesting to say (whereas the converse thesis is false). The second reason is that the doxastic test is spelled out in terms of what one may rationally believe simultaneously, and our pre-theoretical intuitions about that seem to be much more reliable than about the epistemic status of representations.

Firstly, notice that there are many representations that are cognitively distinct, according to the doxastic test, about which the epistemic test has nothing to say. For example, "Cicero is a Roman orator" and "Tully is a Roman orator" are both *a posteriori* even though one could rationally believe one while disbelieving the other.

Similarly, there seems to be *a priori* representations which the doxastic test outputs as

---

[9]This test has received many names in the literature. Schiffer (1978) notoriously dubbed it "Frege's Constraint" on modes of presentation. Schellenberg (2012) calls it Equipollence and bases it on a passage from Frege's 1906 manuscript *Kurze Übersicht meiner logischen Lehren*.

being cognitively distinct about which the epistemic test is completely silent. An easy way to prove that point is by glancing at logical and mathematical statements. Surely, one is not irrational for believing that $p \rightarrow p$ while being unsure whether $\neg\neg((\neg p \rightarrow q) \lor \neg p)$, even though both are *a priori* tautologies. However, I would not want to rely too heavily on that kind of example since it also strikes me as plausible that one would come out as irrational if, *given enough time and calculating skills*, one still held conflicting attitudes towards these statements. That is, when one has conflicting attitudes towards tautologies (or true mathematical statements, e.g. "$2^2=4$" and "$2+2=4$"), this can usually be explained by cognitive limitations on the part of the subject (e.g. the subject was under pressure and miscalculated). Furthermore, these limitations can usually be overcome without recourse to any *a posteriori* information - give him more time, eliminate all distractions, idealize his cognitive capacities etc. Those cognitive limitations are the subject's fault, not the statements', thus I think it is suspicious to claim that rationality assessments are *kosher* in those cases.

In summary, the reason why rational people can hold conflicting attitudes towards *a priori* equivalent representations seem to be totally different from the reason they can hold conflicting attitudes towards cointensional *a posteriori* representations, such as "Cicero is a Roman orator" and "Tully is a Roman orator". In the second case, in opposition to the first, no amount of cognitive idealization would be enough for the subject to change his mind.

Regardless of that, the doxastic case is indeed more comprehensive than the epistemic one, since the former entails that many *a posteriori* statements are cognitively distinct whereas the latter is not even sensitive to their cognitive differences. Furthermore, there seems to be no cases that only the epistemic test is sensitive to.

By itself, this is enough justification to get rid of the epistemic test in favor of the doxastic one. Nevertheless, even if this was not, the mere fact that the epistemic test is defined by means of the technical concepts of *a priority* and *a posterioricity*, while the doxastic relies on what we can rationally believe, should be enough to make us suspicious about its effectiveness. Which definition of *a priority* should we use? Is it "knowledge without recourse to empirical experience"? But, where would this put introspective knowledge? What about reaching the result of a sum by means of an electronic calculator, is this knowledge without recourse to experience? There are many perplexing questions regarding those epistemic concepts, surely it would be a bad omen to start our discussion in such unstable grounds.

Now, even if we decide to go along only with the doxastic test, it is still much less informative than we would like. Notice that it was formulated by means of a material conditional. That means it is only sensitive to statements that one can rationally hold conflicting attitudes towards. What,

then, should we say about those statements that one must hold the same attitude towards? Take a sentence containing an anaphorical pronoun, such as "Cicero was a Roman and he was also an orator". Surely, one would be irrational (or, at least, semantically ignorant) if one held conflicting attitudes to Cicero and the referent of "he". The way anaphorical pronouns work entails that what holds for its antecedent, holds for itself. Still, the doxastic test still does not allow us to say that "Cicero was a Roman" and "He was a Roman" (in which "He" must be read as another instance of the anaphorical pronoun in the previous conjunctive sentence) are cognitively identical.

At this point, one feels slightly tempted to strengthen the doxastic test and rephrase it as a biconditional. That would ensure that any two representations towards which one might not rationally bear conflicting attitudes would be cognitively identical. However, that would be too quick. Take p and $p \rightarrow p,$ for example. It is plausible that no rational subject can bear conflicting attitudes towards these two representations. Their equivalence is, I think, too simple to overlook. Even so, I am not sure whether they are the type of representations that we may want to identify cognitively. Let us bypass those issues for now.

Another limitation of the doxastic test is that it is only sensitive to the cognitive value of complex representations. However, we may also want to claim that tokens may differ in their cognitive values among themselves. Even more, it is very plausible that the cognitive differences between complex representations, such as "Cicero is Cicero" and "Cicero is Tully", are grounded on a cognitive difference between its basic constituents, i.e. "Cicero" and "Tully". Following Dorr (2014), let us call that hypothesis "clausalism":

(Clausalism) All cognitive differences between cointensional representations is due to cognitive differences between its referential tokens.

For a long time, very few philosophers felt the need to argue in favor of (Clausalism), taking it as a given fact that substituting referential expressions or concepts is often enough to change the cognitive value of singular representations. That is based on our pre-theoretical intuition that, somehow, "Cicero" is not a synonym of "Tully". More and more, philosophers are beginning to contest the clausalist hypothesis. A recent contender is Verbalism, according to which the cognitive differences between cointensional representations is due to the attitudes one holds towards them, not the semantic aspects of the representations themselves. Roughly, a verbalist claims that one may rationally bear conflicting attitudes towards the same intensional content given that these attitudes are type-distinct. In other words, one (e.g. Suzanne) may believe that Cicero/Tully is a Roman orator and disbelieve that

Cicero/Tully is a Roman orator given that the types of (dis)belief involved are distinct. The fatal flaw of the verbalist position is that it overrides a believer's disposition of assent and dissent. That is, it completely disregards the fact that Suzanne would describe herself as believing only that Cicero is a Roman orator, but not that Tully also is. Since the verbalist locates cognitive differences in the attitudes one bears towards a coarse-grained intensional content, the difference of referential expressions/concepts that one uses to express those attitudes plays no explanatory role[10]. I intend to adopt a full-blown Clausalism and completely disregard the verbalist hypothesis[11].

If one adopts Clausalism, one has to be able to apply the notion of cognitive value to tokens over and above complex representations. The most natural way to enforce that is by claiming that the cognitive value of complex representations has a constituent structure that reflects the syntactic structure of the representations to which they are associated. In other words, the cognitive value of a complex representation, just like every other ordinary semantic entity associated with complex representations, is compositional. That means it is constituted by the cognitive value of its individual constituents. When two representations differ merely in the substitution of a (referential) token for another, and they have distinct cognitive values, then those two tokens must be associated with distinct cognitive values.

The hypothesis that cognitive value is compositional is, to be sure, a hypothesis. However, as far as hypotheses go, I think this is a very fair one. It would surely be very unfortunate to start our investigation by postulating semantic entities associated with complex representations that are, themselves, not complex (or having a constitution completely independent of the syntax of its owner).

The previous remarks amount to the claim that the cognitive value of a representation is composed by the cognitive values of its syntactic constituents. Since we are only concerned with issues arising from singular representation, the only type of syntactic constituent which we will take the trouble of assigning cognitive values are the referential tokens. However, I do not see any reason against assigning cognitive value to other syntactic constituents, such as predicates or connectives.

Let us now extend the doxastic test so that it gives us information about the cognitive values tokens:

(Tokens Doxastic Test) If two representations differ solely by the substitution of coreferential tokens

---

[10]An even third option would be the hidden-indexical account (Schiffer 1978), according to which the source of cognitive differences is to be located in a third constituent of representations, which has an unarticulated syntactical counterpart.
[11]Bear in mind that Verbalism has nothing to do with the much more familiar thesis that attitudes can be *de re* or *de dicto*. Disregarding Verbalism does not entail disregarding that latter thesis. We will have more to say about the latter in chapter 3.

and one may rationally bear conflicting attitudes towards them, then these coreferential tokens are cognitively distinct.

It should be obvious that cognitive value carves the semantic space much more finely than intensional content (or reference). These new semantic values have a normative dimension: they take into account requirements for a thinker being rational. This normative dimension allows that cognitive value may be used in assessing a thinker's rationality and explaining his behavior. As Loar (1988, p. 99/p. 127) once put it while arguing for cognitive value under the name of "psychological content", they

> individuate beliefs and other propositional attitudes in commonsense psychological explanation, so that they explanatorily interact with each other and with other factors such as perception in familiar ways. […] [psychological content] is that content-like aspect of thoughts, of how we conceive things, by reference to which we consider whether combinations of them are rational, whether they motivate a given belief or action, and so on.

Recently, a couple of philosophers (Almog 2008, Glezakos 2009) have expressed skepticism about the arguments that we have used to ground the postulation of cognitive value over and above intensional content and reference. In the following sections, I will engage their arguments and try to show why they are misguided. In the course of doing that, we will be able to reach a much clearer picture of what cognitive value is and how it is constituted. I will claim that it is based on a particular relation obtaining between certain singular tokens (and not between others): *de jure* coreference.

2.2 WHAT GROUNDS FREGE'S PUZZLE?

The issue of cognitive value is generally discussed by means of identity representations, e.g. "Cicero is Cicero", "Cicero is Tullius" etc. Frege (1948) introduced what would become known as Frege's Puzzle by claiming that identity sentences of the form $a = a$ have distinct cognitive value from sentences of the form $a = b$:

> <a = a> and <a = b> are obviously statements of differing cognitive value; <a = a> holds a priori and, according to Kant, is to be labelled analytic, while statements of the form $a = b$ often contain very valuable extensions of our knowledge and cannot always be established a priori.

The puzzling predicament is precisely that of explaining how identity sentences of those distinct forms can differ in cognitive value when they are identical in reference. Frege's solution was the postulation of Senses, fine-grained semantic values that track cognitive value as opposed to referential content. In that quotation, Frege associates differences in cognitive value with differences in identity sentences' form, but it is never clear which kind of form he is speaking about. Let us try to get clear about that notion.

A familiar answer to this question - though not explicitly stated in *Sinn und Bedeutung* - is that, while sentences of the form $a = b$ contain two distinct referential expressions, a sentence of the form $a = a$ has one referential expression repeated twice over. Under this interpretation, our puzzling predicament becomes that of explaining "what is the source of the cognitive difference between true identity sentences that contain a single name twice, and those that contain two names" (Glezakos, 2009, p. 205).

As innocent as this interpretation might sound, it relies on the not-so-innocent notion of sameness/difference of names. While we seem to have a pre-theoretical grasp on whether some tokens are instances of the same linguistic type or not, things get particularly murky when proper names are concerned (Kaplan, 1990). As Glezakos (2009, p. 206) notes, Frege never settles on a criterion for the difference/identity of proper names that is capable of adequately tracking differences in cognitive value.

An obvious non-starter would be holding that two proper names are the same when they have *the same referent and the same spelling*[12]. Unfortunately, things are not that simple. Notice that even the sentence "Cicero is Cicero" could enable someone to acquire a new belief. This was the general lesson of Kripke's (1979) Paderewski-scenarios[13], where a confused subject encounters the same name on distinct, non-intersecting, occasions, and goes on to think that they refer to distinct individuals. That confused subject mistakenly believes that there are two distinct Roman orators called "Cicero" and learns, upon hearing an utterance of "Cicero is Cicero", that they are, in fact, one.

---

[12]As Glezakos (2009, p. 205) notes, this criterion seems to be implied by Frege's (1948, p. 210) footnote B:

> In the case of an actual proper name such as 'Aristotle' opinions as to the sense may differ. It might, for instance, be taken to be the following: the pupil of Plato and teacher of Alexander the Great. Anybody who does this will attach another sense to the sentence 'Aristotle was born in Stagira' than will a man who takes as the sense of the name: the teacher of Alexander the Great who was born in Stagira. So long as the reference remains the same, such variations of sense may be tolerated […]

[13]A mistaken subject believes that there are two poles called "Paderewski", one, a pianist, the other, a statesman, while in fact she is thinking about the same person who in fact happens to be a pianist and a statesman. Due to his false beliefs, he would not assent to an utterance of "Paderewski is Paderewski". Even though the two tokens of "Paderewski" in that utterance would have the same reference and spelling, they still differ in cognitive value.

In other words, the representation

(**1**) Cicero is Cicero

seems to be simultaneously of the logical form $a = a$ (since it *prima facie* contains the same name twice over) and is informative to the confused subject. If this were the case, "then there would be no epistemic divide between sentences of the form $a = a$ and $a = b$" (Glezakos, 2009, p. 205).

Furthermore, *sameness of spelling* is not even necessary for identity of cognitive value, since any representation containing an anaphorical token, such as:

(**2**) Marcone lives in Venice but he was born in Kiev

contains instances of (type-)distinct expressions that are nonetheless cognitively identical. That is, given that "he" is being read as an anaphorical pronoun (and not a demonstrative), only someone incompetent with the rules of anaphora would bear conflicting attitudes to "Marcone lives in Venice" and "he lives in Venice".

Another obvious non-starter for Frege would be claiming that names are individuated by their associated Senses, i.e. same Sense, same name; distinct Senses, distinct names. Since Frege's Puzzle was originally intended to motivate the postulation of Senses as an inference to the best explanation of a name's cognitive value, presupposing the existence of Senses in its exposition would lead the general Fregean argumentative strategy to circularity and deprive it of any convincing power altogether:

> The puzzle is posed in terms of sentence form, and sentence form is determined by the identity or distinctness of the names appearing in the sentence. If a name is in part individuated by the *Sinn* associated with it, then, in order to be puzzled, one would need to be committed to *Sinne* and their role in determining name identity. If we are not so committed, we will find that the puzzle has no hold of us. (Glezakos 2009, p. 206)

In summary, Glezakos concludes that the logical form of an identity sentence supervenes either (i) on the identity of reference-*cum*-spelling of the expressions therein contained or (ii) on the identity of their associated Senses. If the first strategy is pursued, a sentence of the form $a = a$ comes out as cognitively equivalent to a sentence of the form $a = b$, whereas if the second strategy is pursued,

the puzzle cannot be used as an argument in favor of the postulation of Senses.

There is a lot to be praised in Glezakos' challenge although, as I will demonstrate, its insistence on name identity and logical form is a red herring. To be sure, some responsibility behind this insistence is due to Frege's own decision to associate cognitive value with statements that have the form $a = a$ and $a = b$. On his behalf, notice that Frege's theorizing about "language" was usually aimed at an ideal language deprived of ambiguity as opposed to natural language as we know it[14]. Early philosophers of language, like Frege, Russell and (the early) Wittgenstein, unlike contemporary ones, were notoriously interested in semantics primarily because they believed it could help us to free our thoughts from the imperfections and lack of clarity of natural language, thus enabling us to construct a perfect language, a "conceptual notation" (Frege); a perfect language with which mathematical concepts could be expressed unambiguously.

In view of this agenda, Frege's references to "the language", including the principled distinction between sentences of forms $a = a$ and $a = b$, should be arguably better understood as referring to such an ideal language. Since such a language is completely disambiguated, any two tokens with the same spelling can immediately be known to corefer. If cognitive value were a notion that only applied to ideal languages, it would be enough to define it in terms of spelling. If we were concerned with cognitive value in predicate logic, for example, we could simply identify sameness of spelling with sameness of cognitive value:

(TYPOGRAPHIC COGNITIVE VALUE) Two tokens are cognitively identical (associated to the same Sense) if and only if they are spelled identically

However, natural languages, to which alone Glezakos' challenge applies, are not ideal in that sense. On the contrary, natural languages are so impregnated with ambiguity that typographic facts allow no inference to cognitive facts. Nonetheless, we would still like to say that some identity representations in natural language are trivial whereas others are informative. Now, if natural language's cognitive value is not a matter of linguistic types or typography, what is it a matter of?

I would like to defend that there is a real phenomenon underlying our judgments about cognitive value. That phenomenon is a relation between tokens we shall call "*de jure* coreference".

2.3 DE JURE COREFERENCE

---

[14]Boccardi (2012, p. 135) makes the same point in connection with related issues. Also, Burge (2005, p. 38) and Hahn (1995, p. 174).

The notion of *de jure* coreference (henceforth DJCR) is by now a familiar one in the Philosophy of Language and Mind. This relation between tokens has been given different names in the literature, including, but not limited to, 'strict coreference' (Fine 2007), 'gramatically determined coreference' (Fiengo and May 2006), 'explicit coreference' (Taylor 2015), and '*de jure* codesignation' (Pryor, forthcoming). Other authors have dealt with this relation but did not bother to give it a special name, e.g. Campbell (1987) saw it as a relation that licenses *trading on the identity* of the relevant expressions and Dickie & Rattan (2010) followed him on that. The tag '*de jure* coreference' seems to be gaining more traction in the recent literature so we will use it following the likes of Pinillos (2011) and Recanati (2012, 2015).

In a nutshell, *de jure* coreference tries to capture the relation that is instantiated between two expressions when they represent their content *as the same* in the strongest sense. It is an abstraction from the fact that language and thought enable us to, not only express a certain content, but to re-express it as the same. The re-expression of a content is particularly important to cognition since it is presupposed by such basic capacities as inferential reasoning, memory, the ability to track objects, and communication.

For a more precise definition of DJCR, notice that for any two referential tokens, we seem to have four epistemic possibilities open:

(A) They co-refer

(B) They are both empty

(C) One of them is empty, the other refers

(D) They refer to different particulars

When all of those possibilities are live, we shall say that the two tokens are *independent*, i.e. not DJCR. Think of someone encountering the following utterance for the first time: "Cicero was a Roman and Tully was an orator". It is perfectly reasonable to allow that one may have understood it without having any guess about which of the four referential possibilities are instantiated by "Cicero" and "Tully".

Not all tokens are independent, though. Remember (2), where we had a proper name and an anaphorical pronoun. While one who understands that sentence may wonder whether the expressions are empty – maybe Marcone is just an urban legend – there is no possibility that they will end up either (D) referring to different objects or (C) one of them being empty while the other successfully refers. One would display semantic incompetence by thinking otherwise.

With that in mind, here is a rough and ready definition of DJCR:

(DJCR) Two tokens A and B are DJCR if and only if one who understands them is rationally entitled to believe that they corefer if they refer at all

I intend that rational entitlement be read as some kind of non-evidential justification. That is, if one is rationally entitled to believe that *p*, one is warranted in believing that *p* even in the absence of any evidence in favor of *p*. Taking that into account, notice that DJCR has been defined in terms of what a rational subject is primitively entitled to believe about the reference of certain tokens (that she understands). Equivalently, we could have said that two tokens are DJCR if and only if one is rationally entitled to rule out cases (C) and (D) without recourse to any (empirical) evidence.

DJCR, thus defined, perfectly tracks the intuitions that we used to postulate the doxastic tests a couple of sections ago. When two representations differ solely by the substitution of two DJCR tokens, reason requires us to believe that they have the same truth-value (e.g. it would be irrational to do otherwise). It is easy to see that this requirement entails that one cannot bear conflicting attitudes towards these two representations. Thus, when two representations differ solely by the substitution of DJCR tokens, the doxastic test entails that they are cognitively identical.

The introduction of DJCR has important explanatory benefits. For one thing, it makes it much clearer why it would be irrational to bear conflicting attitudes towards the representations that the doxastic test calls cognitive identical. It would be irrational because these representations contain tokens that are related by DJCR, and that relation makes it irrational to think that they will end up referring to distinct things. They are, so to say, *obviously* about the same thing. DJCR provides a bridge between cognitive value and the coreference relations between tokens.

I take it that DJCR is the relation upon which Frege's talk of an identity sentence's logical form is grounded. That is, an identity representation flanked by independent tokens is precisely that which we will assign the form *a* = *b* whereas an identity representation flanked by DJCR tokens is that we will assign the form *a* = *a*. I will argue that the best version of Frege's Puzzle is posed in terms of DJCR. The puzzling predicament then becomes that of explaining how it is possible that this relation exists in linguistic and mental representation.

2.4 FREGE'S PUZZLE AND SPEAKER-RELATIVITY

Relating DJCR with the cognitive value of identity representations is straightforward: an identity representation flanked by independent tokens is cognitively distinct from a cointensional identity

flanked by DJCR tokens, because it would be irrational to disbelieve the first, but not the former.

These have been the theses defended so far:

(i) Two representations are cognitively distinct when one can rationally bear conflicting attitudes towards them. (Doxastic test)

(ii) Two tokens A and B are DJCR if and only if one who understands them is rationally entitled to believe that they corefer if they refer at all. (DJCR)

(iii) One can rationally bear conflicting attitudes towards two cointensional singular representations when their tokens are independent, i.e. not DJCR.

(iv) [from (i), (ii) and (iii)] Cognitive identity/difference between singular representations supervenes on the instantiation (or lack thereby) of DJCR between its tokens.

(v) [from (iv)] Frege's talk of *the form of an identity* tracks the instantiation (or lack thereby) of DJCR between its relevant tokens;

(vi) [from (v)] Frege's Puzzle is a puzzle about DJCR

Interestingly enough, some of these theses jointly imply the following conclusions:

(vii) [from (ii) and (vi)] Instances of Frege's Puzzle are always related to a particular speaker who understands the relevant representations and is rationally entitled to believe that they corefer if they refer at all.

(viii) [from (vii)] Frege's Puzzle is speaker-relative.

(viii) has, as one of its consequences, that a sentence, such as "Paderewski is Paderewski", has no determinate cognitive value if it is considered independently of a particular speaker who interprets/understands it, i.e. it will be trivial to a speaker who interprets it as stating the self-identity of a Pole, but informative to Kripke's confused subject.

One may then ask, "doesn't this lead us to the conclusion that the cognitive value of an identity representation is an elusive notion that will change according to the idiosyncrasies of every speaker?"

One could argue, for example, that any two linguistic tokens could be interpreted as being DJCR or independent, depending on the idiosyncrasies of their interpreter. That is, a liberal interpreter could take any two tokens and claim that they are, for example, an anaphorical pronoun and its

antecedent. This is certainly possible. Maybe "Aristotle" just is how the anaphorical pronoun is written in the liberal interpreter's native language, so that he would interpret it as being DJCR with a token of "Plato" in "Plato is Greek and Aristotle is a philosopher". If this is the notion of cognitive value that Frege's Puzzle is concerned with, then it is a vacuous puzzle since any two tokens whatsoever can be cognitively identical or distinct in that sense.

It is easy to show that this is not the correct notion to characterize cognitive value for our discussion. Firstly, notice that this sense of cognitive value is a property of token sentences *qua* uninterpreted vehicles of content, e.g. strings of letters or series of sounds, while proper cognitive value is a property of representations *qua* interpreted bearers of content. It is obvious that any linguistic token can be interpreted in any arbitrary way since a string of letters (or a series of sounds) has no intrinsic meaning on its own. DJCR, on the other hand, can only be ascribed to a pair of tokens after one has associated them with meanings. We can then conclude that DJCR, strictly speaking, is not a relation between mere symbols with certain shapes but between tokens interpreted in a determinate way.

Unfortunately, this conclusion seems to imply that Glezakos' threat of circularity has returned through the backdoor. Remember that we are trying to ground Frege's Puzzle on DJCR. Our underlying objective is that this grounding will enable Frege's Puzzle to be advanced – in accordance with Frege's original intention - as an argument in favor of the postulation of Senses, i.e. fine-grained semantic values. However, that objective will only be achieved if DJCR is shown to be conceptually prior to Senses themselves, otherwise what would "emerge is that his ''puzzle'' and his solution are in fact of a piece" (Glezakos, 2009, p. 207). Since we have just concluded that facts about DJCR supervene on facts about tokens' interpretations, it seems that we first need to associate tokens with Senses in order to find out whether they are related by DJCR or not.

I intend to dispel that threat of circularity by showing that DJCR is, indeed, more basic than the assignment of Senses to referential tokens. I will establish that fact through an argument, due to Campbell (1987), demonstrating that DJCR is grounded on the basic cognitive capacity that subjects have of *trading on the identity* of certain tokens in ordinary inferences. Since that capacity does not presuppose any substantial theory of content, it prove that DJCR is explanatorily prior to Senses and may serve to ground their existence.

2.5 THE REGRESS ARGUMENT

Suppose that thinkers were never rationally entitled to believe that some singular tokens are coreferential or not. That is, any beliefs about tokens coreferring with each other would be dependent on evidences (e.g. empirical facts). If something like this were the case, any inference of the following

structure - for "a" meaning a referential expression, "F" and "G" meaning two distinct predicates and the superscripts serving to differentiate between tokens - would come out as being irrational:

I. a^1 is F

II. a^2 is G

III. Thus, something is F and G

The reason why this type of inference would be irrational is that neither its thinker has any primitive reason to believe that "a^1" and "a^2" corefer (if they refer at all) nor the premises of the argument ensure that this fact obtains. If they did not corefer, the conclusion (III) would not be validly inferable. The obvious way to make that a good inference would be adding a further premise containing the missing information, i.e. that the tokens in the first two premises are the same:

I. a^1 is F

II. a^2 is G

III. a^3 is a^4

IV. Thus, something is F and G

It is obvious that this solution would give rise to the same problem as before, for it presupposes that the tokens in the third premise corefer (if they refer at all) with the tokens in the first two premises. However, as we have agreed, the thinker has no entitlement to believe that. The skeptical could always come and ask, "how can you be sure that the third premise is about the same subject matter as the first two?"

Another strategy one could pursue would be to substitute the identity premise (III) for a metalinguistic one, e.g. "a^3" corefers with "a^4". This strategy would fail for related reasons, since the skeptic could question our entitlement for thinking that the expressions in quotes are the same as the ones in the first two premises[15]. Finally, one could substitute the identity premise (III) for a descriptive premise, e.g. the H = the J, for some H that is uniquely satisfied by the token in (I) and some J that is uniquely satisfied by the token in (II). That proposal would also fail to elude the skeptic since one would still need to add the premises *a^3 is the H* and *a^4 is the J* in order to derive the unifying

---

[15] Interestingly, that observation seems to entail that there can be a relation similar to DJCR between a use and a mention of an expression. That is, in "The proper name 'John' refers to John", the mention and the use of "John" seem to be more closely related than their analogues in "The proper name 'Mr. Smith' refers to John".

conclusion, and the skeptic would simply point out that we have no guarantee of these tokens' coreferring with the previous ones.

One could produce as many tokens as one desired, still, one's predicament would be the same: the inference would anyway come out as irrational due to the lack of some basic entitlement. The possibility of reasoning correctly with referential expressions would, consequently, be led into a regress.

The moral that Campbell (1987, p. 276) influentially extracted from these observations was that "we need an account of when an inference may simply trade upon the coreference of two singular tokens" if the possibility of rational singular reasoning is to be established at all. In other words, thinkers must, at some point, be able to treat distinct tokens as being univocal without need of any additional information.

Even more so, that moral has to be taken into account by any theory of singular content that takes seriously the idea that some inferences are rational whereas others are not – and that is such a basic presupposition that it would seem to affect any theory of content *tout court*.

It is important to notice that even a referentialist theory of content, according to which singular content is (at most) *intensional content*, will need to have something to say about our entitlement to *trade on the identity* of some distinct referential tokens. Although these theorists will insist that we never have *a priori* knowledge about any two tokens corefererring, they will need to provide a way of explaining why we are sometimes entitled to simply assume that some tokens corefer, if they refer at all.

Here is a very vivid way to summarize the conclusion of the regress argument: if rationality never provided thinkers with (non-evidential) entitlement to believe that some tokens corefer (if they refer at all), there would be no normative/epistemic/rational difference between a subject A who infers:

I. Hesperus is shiny

II. Hesperus is round

III.Thus, something is both shiny and round

and a subject B who infers:

I. Hesperus is shiny

II. Phosphorus is round

III.Thus, something is both shiny and round

Assuming that rationality never provides thinkers with entitlement about two tokens coreferring, the premises of the first argument might be about different subject matter just as much as the premises in the second one. I gather that this is absurd enough to be considered a *reductio*. That is, it seems obvious that A must be deemed rational (whereas B should be deemed irrational). Furthermore, this fact is completely independent of the details of one's preferred solution to Frege's Puzzle (or the details of one's semantic theory).

Gerken (2011, 2013), for instance, is a paradigmatically referentialist philosopher that still goes to great lengths in order to explain how some cointensional inferences can be rationally different, although he denies that this difference tracks any semantic fact. According to him, an inference is rational when it is reliably governed by a cognitive (non-semantic) competence, *the univocality competence*. This author defends the claim that rational reasoning does not supervene on classical validity, so that some inferences can be simultaneously irrational but valid (or simultaneously rational but invalid). Similarly, Faria (2009) claims that some invalid arguments are safe - rationally excusable - while others are unsafe - rationally blameworthy – although the difference between them amount only to a matter of degree.

In summary, our entitlement to *trade on the identity* of certain referential tokens is a *datum* for any theory that distinguishes between normatively good and bad singular inferences.

At this point, we can introduce the concept of DJCR as the relation that tokens instantiate when we are entitled to *trade on their identity*. It can now be seen that DJCR is merely an abstraction of a general normative fact that any reasonable theory of content must make sense of. As opposed to presupposing a particular theory of content, e.g. a Fregean theory of Senses, the acknowledgment of the DJCR relation between singular tokens is a precondition on the reasonableness of any putative theory.

With that in mind, we can finally restate Frege's Puzzle in a way that, I hope, evades the threat of vacuity or circularity. After taking the previous discussion into account, the Puzzle becomes that of explaining how we are able to *trade on the identity* of certain tokens. In other words, our puzzling predicament becomes that of explaining in virtue of what some tokens are DJCR while others are not:

(Frege's Puzzle) In virtue of what are some tokens DJCR (such that their identity is trivial) while others are independent (such that their identity is informative)?

Of course one could respond to that by offering a non-semantic explanation of DJCR. This

is indeed the line of response of many referentialists (e.g. Salmon 1986, Soames 2002, Wettstein 1989, Gerken 2013, Faria 2009). However, this is not Glezakos' line of attack, which tries to undermine the mere possibility for Frege to set up the puzzle in non-circular terms.

Frege's own reply to the puzzle was, of course, the postulation of Senses, i.e. semantic values which track doxastic aspects. The obvious way to update his solution in light of our new exposition of the puzzle is by claiming that (i) tokens are DJCR when they have the same Sense and (ii) that tokens are independent when they have distinct Senses. The explanation of DJCR through *sameness of Sense* is both theoretically cogent and exegetically coherent.

2.6 TAKING STOCK

Let's take stock of the main findings from the last five subsections. Frege's discussion takes off from the fact that cointensional representations may differ in their cognitive value, where cognitive value was individuated by the doxastic test (for complex representations and for tokens). Subsequently, I have argued that the doxastic test should be seen as emanating from the concept of DJCR, which I have thus defined:

(DJCR) Two tokens A and B are DJCR if and only if one who understands them is rationally entitled (on the sole basis of this understanding) to assume that they corefer if they refer at all

The obvious way to connect cognitive value and DJCR was, we have seen, to identify the cognitive identity of two tokens with them being DJCR with each other. This also works the other way round: when two tokens are independent, they are cognitively distinct:

(A2) Two tokens A and B are cognitively identical if and only if A and B are DJCR.

Furthermore, I have showed that DJCR is grounded on the cognitive capacity that thinkers have of *trading on the identity* of some referential tokens. Since that capacity is a basic *datum* for a theory of content, posing Frege's Puzzle in terms of DJCR does not fall prey to Glezakos' (2009) claim of conceptual circularity.

Finally, we have suggested that, if one wants a semantic solution to the puzzle (as the acceptance of (A1) would require), one natural way to go is to follow Frege in postulating Senses, fine-grained semantic values, and claiming that *sameness of Sense* between two tokens explains them being DJCR with each other:

(A3) Two tokens A and B have the same Sense if and only if A and B are DJCR.

From (A2) and (A3) we can deduct:

(A4) Two tokens A and B are cognitively identical if and only if A and B have the same Sense

Interestingly, this way of framing Frege's argument in favor of Senses reminds us closely of his use of abstraction principles to argue for the existence of numbers (Frege, 1953). The abstraction principle asserts that:

(Arithmetical Abstraction Principle) The number of Fs = the number of Gs iff Fs and Gs are related by the equinumerosity relation[16].

The obtaining of the equivalence relation on the right hand side of the biconditional, he argued, ensures that the expressions flanking the identity sign are not empty, i.e. that numbers exist. Likewise, Frege seems to have had the following principle in mind in his master argument for the existence of Senses:

(Senses Abstraction Principle) The sense of A = the sense of B iff A and B are related by the DJCR relation.

Essential to his reasoning in the case of numbers is the fact that equinumerosity is provably an equivalence relation: it is reflexive, symmetric and transitive. DJCR, however, unlike equinumerosity, can be argued to be not an equivalence relation, and this constitutes a potential lethal objection to Frege's solution, even under the light in which we have put it.

In the next chapter we will take a look at a set of objections coming from Fine (2007) and Pinillos (2011) intending to demonstrate that DJCR is a non-transitive relation. If that were the case, DJCR would not be able to ground the abstraction of particular entities, so that the postulation of Senses would fail a solution to Frege's Puzzle. Furthermore, we will assess whether the postulation of Senses is theoretically interesting even if the non-transitivity arguments can be met. After all, why

---

[16]This principle is known as "Hume's Principle".

should one ever postulate abstract entities instead of sticking to an equivalence relation?

3. INTRINSICALISM *versus* RELATIONISM: ON TRANSITIVITY

In the last chapter, we saw that Frege's Puzzle is fundamentally an issue about a relation, DJCR, which rationally entitles thinkers to believe that some distinct tokens corefer (if they refer at all). The puzzling predicament was then shown to be that of explaining the grounds of that relation. We showed that this predicament affects not only those who defend (A1), i.e. that cognitive value must be explained semantically, but, on the contrary, any philosopher who takes seriously the idea that some singular inferences are rational and some are irrational.

In this chapter, we will reinforce (A1) and simply dismiss any solution to Frege's Puzzle that is not semantic. Our objective will be assessing, among the semantic solutions to the puzzle, which one is the best.

Looking from a very abstract point of view, there seems to be two kinds of solution to the puzzle. Following Fine (2007), let us divide the logical space of possible semantic solutions to the puzzle into *intrinsicalist* and *relationist* theories:

(Intrinsicalism) If A and B are DJCR then A and B are each associated with some hyper-intensional content which is identical.[17]

(Relationism) If A and B are DJCR then there is a semantic relation occurring between A and B even though their individual semantics is exhausted by their referents.

The adoption of (Intrinsicalism) entails that DJCR is grounded on a certain (hyper-intensional) content possessed by individual tokens. Thus, since a Fregean theory explains DJCR by the possession of identical Senses by a pair of tokens, it is an example of an intrinsicalist theory.

Adopting (Relationism), on the other hand, enables one to maintain that two DJCR tokens may, nonetheless, be semantically identical. In other words, the relationist philosopher locates the hyper-intensionality of singular representation in a semantic level that is extrinsic to the semantic contents of the individual tokens. That means that a relationist philosopher can stick with the thesis that the semantic content of a token is (at most) intensional content, while still providing an explanation of the difference between cointensional DJCR and cointensional independent tokens.

It is interesting to consider the distinction between Intrinsicalism and Relationism in the light of our previous discussion about Principles of Abstraction. Our reconstruction of Frege's master

---

[17]Fine's (2007, p. 39) own definition of Intrinsicality is: "If the pairs of names 'Cicero', 'Cicero' and 'Cicero', 'Tully' are semantically different, then so are the names 'Cicero' and 'Tully'."

argument for the postulation of Senses entailed that, at some point, he made use of a Principle that abstracted Senses out of the relation of DJCR. One can then generalize and define a theory as Intrinsicalist if and only if it assumes an instance of the Semantic Principle of Abstraction:

(Semantic Principle of Abstraction) Two tokens A and B are related by DJCR if and only if there is an entity, i.e. a hyper-intensional semantic content, associated with both A and B.

Relationism, on the other hand, can be understood as the refusal to take the road towards abstraction. If one can explain all the relevant semantic *data* with a primitive relation, without postulating a set of entities to serve as individual tokens' semantic values, why should one insist on abstraction? In other words, there is nothing mandatory about taking DJCR to be grounded on individual tokens' semantic values instead of grounded on a primitive semantic relation that is over and above their individual semantic values.

At this point, there are two interesting, but distinct, questions one could pose. First, is there any *desideratum* that could only be explained by one type of theory but not the other? That is, do these two types of theory diverge in their explanatory potential only because one accepts, while the other rejects, the Semantic Abstraction Principle? Second, even if the best intrinsicalist theory is explanatorily equivalent to the best relationist theory, are there any reasons to prefer one in detriment to the other? In other words, is the Semantic Principle of Abstraction desirable even if it leads to no gain/or loss in semantic explanation?

In this chapter, we will deal with both questions in relation to a couple of arguments for the non-transitivity of DJCR (Pinillos 2011): the anaphorical case (3.1) and the attitudinal case (3.2). These arguments purport to stablish that no intrinsicalist theory can correctly explain DJCR, since the latter is a non-transitive relation, whereas every intrinsicalist theory is committed to identifying it with an equivalence relation. If these objections are to the point, they constitute a transcendental argument against any intrinsicalist theory. I will claim that both arguments can be rebutted by the intrinsicalist, although we will have to improve our – so far minimal – understanding of what properties cognitive values must possess.

3.1 THE ANAPHORICAL CASE

Consider this example from Pinillos (2011, p. 6):

(3) We were debating whether to investigate both **Hesperus** and **Phosphorus**; but when we got

evidence of their true identity, we immediately sent probes **there**.

It seems undebatable that 'Hesperus' and 'Phosphorus' are as independent as two tokens can be. For one thing, one could understand (3) without having any opinion about whether they corefer. On the other hand, both tokens seem to be DJCR with the anaphorical pronoun 'there': it is not open to an interpreter to doubt whether 'there' corefers with both 'Hesperus' and 'Phosphorus' – that would amount to failing to understand how an anaphorical pronoun works.

Let us assume that the Fregean theory of Senses is the best intrinsicalist theory available. That means that the best intrinsicalist explanation for DJCR is by means of the relation of *sameness of Sense*. Now the challenge is set up: if DJCR is to be explained by the identity of Sense, 'there' would come out as having the same Sense as both 'Hesperus' and 'Phosphorus' even though their ['Hesperus' and 'Phosphorus''s] Senses are distinct. Pinillos takes (3) to show that DJCR is non-transitive, since an expression token – 'there' – is DJCR with two others – 'Hesperus' and 'Phosphorus' – even though these two are not DJCR with one another, i.e. 'Hesperus' is DJCR with 'there'; 'there' is DJCR with 'Phosphorus' but 'Phosphorus' is not DJCR with 'Hesperus'. Identity (of Senses), on the other hand, is an equivalence relation, so it is necessarily transitive. The conclusion is that no relation of identity can be used to explain a non-transitive relation such as DJCR, thus, (Intrinsicalism) is false. What could the Fregean reply?

Firstly, let us try to spell out what is going on in (3) according to our pre-theoretic semantic intuitions. It seems fair to say that the linguistic rules governing an anaphorical pronoun - such as 'he' in

(2) Marcone lives in Venice but he was born in Kiev

entail that it must refer to whatever is the referent of its antecedent expression – thus, 'he' must refer to whatever 'Marcone' refers to. An anaphorical pronoun seems to be simply a proxy for another expression that is doing all the heavy semantical labor.

(3) introduces a higher degree of complexity than (2) because its anaphorical expression, 'there', is not connected to only one antecedent, but to two. Extrapolating from our proto-theory of anaphora, we could say that an anaphorical pronoun connected to multiple antecedents must refer to the one object all of its antecedents refer to. If this is the correct rule governing multiple-antecedents anaphora, 'there' must refer to whatever is the referent of both 'Hesperus' and 'Phosphorus'.

While this sounds fair, notice that it is enough to show that the relation instantiated between

'there' and its antecedents is distinct from DJCR as we have previously defined it. It is possible that 'there' fails to refer while 'Hesperus' (or 'Phosphorus') successfully refers. That is possible because 'there' must refer to the one object all of its antecedents refer to, but, since 'Hesperus' and 'Phosphorus' could, for all that a speaker is rationally entitled to believe, refer to distinct objects, 'there' could reasonably fail to pick out an univocal referent and come out empty. That would be true if, for example, 'Hesperus' referred to Venus while 'Phosphorus' referred to Jupiter; while this is not actually the case, it is not something speakers are rationally entitled to rule out. In summary: one is not entitled to rule out that Case (C) obtains between the pairs 'there'/'Hesperus' and 'there'/'Phosphorus'. Since DJCR, *per* definition, entitles thinkers to rule out Case (C), we seem to be observing a closely related, but distinct, relation of coreference between expression tokens.

Following Recanati (2015, p. 16), let us call this distinct relation weak *de jure* coreference (W-DJCR):

(W-DJCR) Two tokens A and B are W-DJCR if and only if one who understands them is rationally entitled to believe that either they corefer (if they refer at all) or that one of them refers while the other does not.

Alternatively, two tokens are W-DJCR if one is rationally entitled to rule out Case (D), i.e. that they successfully refer to distinct objects.[18]

It follows that DJCR entails W-DJCR, but not the converse. If one is entitled to believe that some tokens A and B corefer (if they refer at all), then one is entitled to believe that they do not refer to distinct objects. On the other hand, if one is only entitled to believe that two tokens A and B do not refer to distinct objects, one is not necessarily entitled to believe that they corefer (if they refer at all),

---

[18]Other two putative cases of W-DJCR in natural language include:

(i) Semantic Deference to Multiple Individuals: Capellen (2012) argues that the word 'intuition' has been defined in multiple inconsistent ways by different philosophers/psychologists. If someone decided to use that word deferring its meanings to the experts of its community (there is nothing stopping you or me from doing that!), one would end up having producing tokens of 'intuition' that are W-DJCR with the tokens of 'intuition' produced by the disagreeing philosophers/psychologists, but not DJCR.

(ii) Composite Descriptive Names: it is plausible that descriptive names, i.e. directly referential names that have their reference fixed by a definite description, exist in natural language. Some examples would include "Jack the Ripper" (its associated description being something like "the unique killer of countless women in Victorian London") and "Deep Throat" ("the unique informant of journalists during the Watergate scandal"). If that much is true, there is nothing stopping me and you from introducing a new descriptive name in our language which would have its reference fixed both by the description of "Jack the Ripper" and that of "Deep Throat". That is, we could introduce the descriptive name "Deep Throat the Ripper" and stipulate that it refers to someone if and only if this person is the unique killer of countless women in Victorian London and the unique informant of journalists during the Watergate scandal. "Deep Throat the Ripper" would then be W-DJCR with both "Jack the Ripper" and "Deep Throat", but not DJCR.

since it would still be rational to wonder whether one of them successfully refers while the other does not.

In summary, every pair of expressions having the property of DJCR also has the property of W-DJCR, but some pairs of expressions, e.g. 'there' and 'Hesperus' (or 'Phosphorus') in (3), have W-DJCR, but not DJCR.

If anaphorical pronouns with multiple antecedents come out empty when their antecedents do not corefer, W-DJCR has true instances in natural languages such as English and is not merely a technicality. That being the case, it is to be expected that any theory concerned with the cognitive value of expression tokens will provide an explanation for it. My main point in the next few subsections will be that the cognitive import of two tokens which are merely W-DJCR is different from that of two tokens which are DJCR, so that any theory which fails to draw the distinction between these two relations will be glossing over important *data*.

On one hand, we agree with Pinillos (2011) that, *contra* Frege, sameness or difference of Fregean Senses will not do that job, since W-DJCR really is a non-transitive relation. On the other hand, we disagree that sameness or difference of Senses is the only tool a Fregean theory has to explain the phenomena associated with cognitive value.

To expand on that point, we need to be more explicit about what properties Senses must possess.

3.1.1   SENSES AS CONDITIONS ON REFERENCE

Up until this point, I have worked with a very deflated understanding of Senses according to which they are just those entities which track differences of cognitive value (while cognitive value was explained in terms of DJCR). If Senses are to explain how speakers are rationally entitled to believe that two tokens corefer (if they refer at all), then they must have some bearing on the reference conditions of the tokens which they are associated to.

I take this to suggest that Senses can be minimally understood as *conditions an object has to satisfy in order to be the referent of a certain token*. I intend that this thesis be read in a very minimal sense, since it does not commit me to any substantial theory of what Senses are. That is, defending that Senses are *conditions* does not weigh in favor of any specific Fregean theory of Senses, e.g. descriptivism, singularism (Recanati 2012), epistemic bidimensionalism (Chalmers 2011) etc. Although, to be sure, it does commit me to a theory of Senses in which they are, to use a Dummettian expression, 'routes to a referent', that is, cognitive paths that provide a means to discover the referent of a token. It is important to notice that this thesis was not at play up until this point and that it is being

introduced with a specific purpose in mind, i.e. overcoming Pinillos' challenge.

Continuing, since *conditions* are the kind of things that can be aggregated to form composite entities, this thesis entails that Senses also can, i.e. that they can enter into composition and decomposition relations. This means that it is meaningful to speak of a certain Sense being composed by two other Senses and/or of a certain Sense being a part of another:

(A5) Any two Senses α and β can be composed to form the composite Sense α *and β*[19]

The acceptance of (A5) is enough to sketch a Fregean explanation of the reference relations in (3): let α be the Sense of 'Hesperus' and β, that of 'Phosphorus'. Because the condition for being the referent of 'there' is that it must also be the referent of 'Hesperus' and 'Phosphorus', its Sense must be constituted by the conditions prescribed by the Senses of both expressions. The obvious way to flesh out that conclusion is by claiming that the Sense of 'there' is composite, taking both α and β as its parts.

Thus, while the condition that an object has to satisfy in order to be the referent of its antecedents would be that prescribed by α, for 'Hesperus', and by β, for 'Phosphorus', the condition associated with 'there' would be the composite one prescribed jointly by α and β.

Since an object can satisfy a certain condition without satisfying a more complex condition that is partially constituted by the first one, W-DJCR is explained as the relation that obtains between two expression tokens A and B when their Senses are constituted by a part in common[20]:

(A6) Two tokens A and B are W-DJCR if and only if either the Sense of A has a part in common with the Sense of B or the Sense of B has a part in common to the Sense of A.

Since every Sense has a part in common with itself, DJCR (i.e. identity of Senses) entails W-DJCR. On the other hand, some Senses sharing a part are not identical, thus, W-DJCR does not entail DJCR.

To be sure, our intrinsicalist reply to Pinillos' argument depends on the acceptance of (A5). Thus, one moral that we can already take from the previous discussion is that: if (Intrinsicalism) is true, the hyper-intensional contents it postulates must be capable of instantiating relations that are more refined than identity or difference.

---

[19] I am assuming only singular tokens are associated to Senses, i.e. that every Sense is singular. Otherwise, (A5) would entail the weird conclusion that one my compose Senses of distinct orders, such as the Sense of a logical connective with the Sense of a predicate.

[20] The way I am using "a part of a Sense" entails that every part of a Sense must itself be a Sense on its own.

I will not dwell on this point for more than one paragraph, but notice that a syntactic solution to Frege's Puzzle à là Fodor (2008, chapter 3), in which it is claimed that DJCR is to be explained by identity of syntactic symbol in the Language of Thought, would be a kind of "intrinsicalist" solution to the puzzle[21] that provides no clear way to accommodate (A5). That is, simple syntactic symbols (think of individual constants) are not the kind of stuff that enter into composition relations.

I will conclude that section by reversing the table on Pinillos' challenge. That is, whereas he advanced the non-transitivity argument as a reason to reject (Intrinsicalism) in favor of (Relationism), I will show why an intrinsicalist theory that accepts (A5) and (A6) is better equipped to explain the phenomena of cognitive value than its relationist rival.

3.1.2   REVERSING THE TABLES ON RELATIONISM

A relationist theory, as we have seen, denies that there is any intrinsic semantic difference between coreferential tokens that differ in cognitive value. Their cognitive discrepancy is wholly explained by the presence (or absence) of a primitive semantic relation linking them which is over and above these tokens' intrinsic semantic values.

Let's take a real example of relationist theory from the literature. Pinillos' (2011) particular brand of relationist semantics calls its main semantic relation "P-Linking" and defines it as a relation obtaining between two tokens if and only if they are coreferential if they *both* refer at all (Pinillos, 2011, p. 28).

Thus defined, P-Linking is equivalent to the relation of W-DJCR[22]. This can be proved by noticing that

(P-Linking) Two tokens A and B are P-Linked if and only if they corefer if they both successfully refer

entails that, if two tokens A and B are P-Linked, then one is rationally entitled to believe that A and B either (i) corefer, (ii) that they are both empty or (iii) that one of them refers while the other does not[23].

---

[21]Strictly speaking, we have defined "intrinsicalism" to encompass only semantic solutions to the puzzle. Nonetheless, I expect that my general point can be understood.

[22]I am obviously glossing over the absence of any mention to a speaker or rational entitlement in Pinillos' definition of P-Linking. These facts are not important to the current discussion.

[23]If one defined (P-Linking) and (W-DJCR) in terms of the four cases of coreference, this is what we would get:

   (Case P-Linking) Two tokens A and B are P-Linked if and only if: if Case 2 is not the case and Case 3 is not the

However, as we have showed, tokens can instantiate two distinct relations of coreference, DJCR and W-DJCR. It is easy to see that Pinillos' P-Linking does not have enough granularity to track the cognitive import that is peculiar to each of these two relations. That follows immediately from the fact that P-Linking is, *per* definition, equivalent to W-DJCR, but not equivalent to the stronger DJCR. Notice that the two DJCR tokens in (2) satisfy the P-Linking axiom just as much as the two tokens that are merely W-DJCR in (3):

(2) Marcone lives in Venice but he was born in Kiev.

(3) We were debating whether to investigate both **Hesperus** and **Phosphorus**; but when we got evidence of their true identity, we immediately sent probes **there**.

Even though "Marcone" and "he" are related by DJCR whereas "Hesperus" and "there" are related solely by W-DJCR, these two pairs of tokens equally satisfy Pinillos' axiom, i.e. the two pairs would corefer if *every* token successfully referred. This means that Pinillos' theory - although it has the virtue of introducing a non-transitive relation - makes hush of the distinction between the two types of coreference relations we have unveiled. If that is the case, his theory is silent about there being any cognitive differences between DJCR tokens and (merely) W-DJCR ones.

However, it is easy to prove that there are significant cognitive differences between those relations. What follows is an argument to that effect.

Imagine that one learns that the referent of "Marcone" in (2) is a cook. One could then immediately conclude that the referent of "he" in (2) also is a cook. This also works the other way around, if one had first learned that the referent of "he" is a cook, one could have immediately

---

case, then Case 1 is the case.

    (Case W-DJCR) Two tokens A and B are W-DJCR if and only if: if Case 2 is not the case, then Case 1 is the case or Case 3 is the case.

Now, take "p" for "Case 2 is the case"; "q" for "Case 3 is the case" and "r" for "Case 1 is the case".

It can now be easily seen that the form of P-Linking is:
If not-p and not-q, then r,

while the form of W-DJCR is:
If not-p, then q or r.

A simple proof is enough to establish that they are logically equivalent (I leave that as an exercise to the reader).

Just for the sake of completeness, DJCR has the logical form:
If not-p, then r.

concluded that the referent of "Marcone" also is. In summary, when two tokens are DJCR, one can deduct freely from one to the other and vice-versa.

Observe how things are different when the (merely) W-DJCR tokens in (3) are concerned. If one learns that the referent of "there" is the second planet from the Sun, one can immediately conclude that the referent of "Hesperus" also is. However, the converse does not hold: if one had first learned that the referent of "Hesperus" is the second planet from the Sun, one could not have concluded immediately that the referent of "there" also is. W-DJCR tokens allow that a Case 3 obtains (i.e. one expression refers while the other does not), so one cannot infer directly from "Hesperus" having a referent that "there" will also have one.

In more precise terms, when there is DJCR between any two tokens A and B, we are entitled to make these two schematic inferences (let us call them "substitution inferences") where F(A) and F(B) differ solely by the substitution of A for B:

$$F(A) \vDash F(B)$$
$$F(B) \vDash F(A)$$

However, when A and B are merely W-DJCR, there is no guarantee that *either* substitution inference will be entitled. We have already seen that, as far as the tokens in (3) are concerned, W-DJCR seems to license one of the substitution inferences (going from "there is F" to "Hesperus is F") but not the other (going from "Hesperus is F" to "there is F"). Even more than that, there are cases where both substitution inferences are irrational.

I will now present a representation that is an expanded version of (3). My objective will be showing that there can be (merely) W-DJCR tokens which do not license *any* of the two substitution inferences. More generally, I intend to make explicit the cognitive difference between DJCR and W-DJCR tokens by defending the following thesis:

(**Cognitive difference between DJCR and W-DJCR tokens**) For any two DJCR tokens, but never for merely W-DJCR tokens, we are entitled to make both substitution inferences with them.

Consider this expanded representation based on (3):

(**4**) We were debating whether to investigate both **Hesperus** and **Phosphorus**[1]; but when we got evidence of their true identity, we immediately sent probes **there**. Since that was also evidence to the

45

effect that **Venus** just is identical to **Phosphorus[2]**, researching **that planet** became our most urgent priority.

(4) has two parts. The first is a perfect replication of (3), where an anaphorical pronoun ("there") is related to two independent antecedents ("Hesperus", "Phosphorus[1]"). The second part of (4) contains three tokens that instantiate the same relations as those in the first part, that is, it also contains an anaphorical pronoun ("that planet") which is related to two independent antecedents ("Venus", "Phosphorus[2]"). Finally, the most natural reading of (4) entails that "Phosphorus[1]" and "Phosphorus[2]" are DJCR, since they would be taken to be two instances of the same name. Therefore, the conditions for being the referent of "Phosphorus[1]" are identical to the conditions for being the referent of "Phosphorus[2]".

If this is correct, we can infer that "there" could never refer to anything different from "that planet". That is so because the conditions for being the referent of "there" have a part in common to the conditions for being the referent of "that planet", i.e. being the referent of "Phosphorus[1]"/"Phosphorus[2]". Nonetheless, it is possible that (i) "there" fails to refer while "that planet" successfully refers or that (ii) "that planet" fails to refer while "there" successfully refer. (i) would be true if "Hesperus" were the only token in (4) that failed to refer, whereas (ii) would be true if "Venus" were the only token in (4) that failed to refer.

Summarizing: "there" and "that planet" are, *per* definition, W-DJCR. That is, it is impossible that they end up referring to distinct objects (but it is possible that one is empty while the other is not). On the other hand, it is easy to see that none of the substitution inferences containing them are rationally licensed:

F(there) $\nvDash$ F(that planet)

F(that planet) $\nvDash$ F(there)

Since it could be the case that "there is F" is true while "that planet" fails to refer, one may not make the first inference. Conversely, it could be the case that "that planet is F" is true while "there" fails to refer, so one may also not make the second inference. Thus, W-DJCR between two tokens A and B does not necessarily license any of the substitution inferences.

From these points, I conclude that there is a significant cognitive difference between tokens which are DJCR and tokens which are merely W-DJCR. DJCR is clearly the stronger relation of the two, since it entitles us to treat its *relata* as *samesaying* in a very strong sense (whatever one believes of

the referent of a certain token, one **must** also believe of the referent of the tokens that are DJCR to it). W-DJCR, on the other hand, is considerably weaker since it does not guarantee entitlement to any of the substitution inferences. We have observed the tokens in (3), where only one of the inferences was licensed, and the tokens in (4), where none of the inferences was licensed. W-DJCR is an intermediary relation between *samesaying* and complete *independence*.

Furthermore, I argue that these cognitive differences stem from the same source as those differences in the more traditional Frege-cases. For that reason, they should receive a unified explanation. I will now show that the relationist theory does not have the resources to offer that explanation, as opposed to its Fregean rival.

First, let us look at the Fregean explanation. Since that theorist explains DJCR between tokens by means of their associated Senses, where Senses are minimally understood as the conditions for being the referent of a certain token, she will explain the validity of substitutions inferences by means of the Senses contained in the premises and conclusion. When two tokens A and B are DJCR, the Fregean postulates a certain Sense α which is associated with A and with B. Since the conditions for being the referent of A and B are identical, it is impossible that they refer to distinct objects (of that one is empty while the other is not). Any thinker who knows all of this will be entitled to make any of the two substitution inferences with A and B.

As we have seen, W-DJCR between tokens A and B is explained, by the Fregean, by means of Senses that have a part in common. In (3), the most natural explanation was that "there" has a composite Sense *α and β* which is composed by the individual Senses of its two antecedents. The reason why only one of the substitution inferences are valid in this case is that, if the conditions prescribed by *α and β* are satisfied by a certain object, one can infer that the conditions prescribed by *α* also are, but not the converse.

In (4), on the other hand, we seem to have two distinct composite Senses: that of "there" and that of "that planet". While the Sense of "there" is composed by the Senses of "Hesperus" and "Phosphorus[1]", the Sense of "that planet" is composed by the Senses of "Phosphorus[2]" and "Venus". Thus, these composite Senses both have a different part (the Senses of "Hesperus" and "Venus") and a part in common (the identical Sense of "Phosphorus[1]" and "Phosphorus[2]"). Taking that into account, we could schematically represent the Sense of "there" as being *α and β* and the Sense of "that planet" as *β and δ*. It is now easy to see why none of the substitution inferences are licensed: it is possible that *α and β* is successfully satisfied by an object while *β and δ* fails to be, and vice-versa.

In conclusion, the Fregean (with the aid of (A5)) provides a unified explanation, by means of the introduction of Senses, for why substitution inferences are valid for DJCR tokens but not

necessarily valid for W-DJCR tokens.

I take that these facts about substitution inferences are facts about the cognitive value of tokens, since what we are rationally entitled to infer using referential expressions is part of the doxastic/epistemic characteristics of representations. Thus, the Fregean has a unified explanation about tokens' cognitive values which highlights the distinction between DJCR and W-DJCR.

What about a relationist such as Pinillos, what could he offer to explain those cognitive differences? As we have seen, the relationist theory is a one trick pony: it sets out to characterize all relevant cognitive phenomena by means of one primitive semantic relation (P-Linking, in the case of Pinillos (2011)). But P-Linking is a relation that is equally instantiated by DJCR and W-DJCR tokens, so it does not suffice to draw any cognitive distinction between them.

One obvious way to tweak Pinillos' theory would be adding a distinct primitive semantic relation, e.g. Z-Linking, that exclusively tracked DJCR:

(Z-Linking) Two tokens A and B are Z-Linked if and only if they corefer if *at least one of them* successfully refers

Pinillos could then go on to identify Z-Linking with DJCR and P-Linking with W-DJCR. The cognitive difference between DJCR tokens and W-DJCR ones would then be explained by the instantiation of a distinct primitive relation, Z-Linking, in the first case, but not in the second.

Nonetheless, this tweaked theory would be inferior to its rival, the Fregean, for two main reasons.

Firstly, it postulates two primitive semantic relations to explain phenomena that are deeply interrelated; that DJCR and W-DJCR are defined in such similar terms makes us expect that they receive a uniform explanation. However, Pinillos' tweaked theory makes them come out as wholly independent, i.e. based on distinct primitive relations.

Secondly, even a Bi-Relationist theory, i.e. a relationist theory with two primitive relations, would not suffice to explain why **both** substitution inferences fail for some W-DJCR pairs of tokens (such as "there" and "that planet") as opposed to **only one of them** for some other W-DJCR pairs of tokens (such as "there" and "Phosphorus"). Since both pairs equally instantiate P-Linking, there is nothing in that relationist theory carving a distinction between them. The obvious strategy to overcome that shortcoming would be adding even more primitive semantic relations to distinguish between these two kinds of cases. It strikes me as obvious that this would make Pinillos' theory even less parsimonious and undesirable.

In conclusion, explaining DJCR and W-DJCR by means of *sameness of Sense* and *Senses that have a part in common* enables the explanation of multiple cognitive phenomena associated with coreference, such as why some substitution inferences are always valid in cases of DJCR, but not necessarily valid for cases of W-DJCR. The relationist explanation of DJCR, on the other hand, was shown to suffer from a structural limitation: since it postulates only one primitive semantic relation, there is no obvious way to distinguish between DJCR and W-DJCR. We have showed one way to tweak the relationist theory by adding a second primitive relation, so that one has as many semantic relations as coreference relations. However, even this tweaked theory would still fail for two reasons: (i) it would be ontologically less parsimonious than the Fregean theory and (ii) it still would not provide us with any explanation of why some W-DJCR tokens have different cognitive value in relation to substitution inferences.

3.2 THE ATTITUDINAL CASE

Consider the following case from Pinillos (2011):

(5) Smith[1] is wearing a costume, and (as a result) Sally thinks he is someone other than Smith[2].

Pinillos takes this representation to provide an additional argument in favor of the non-transitivity of DJCR. His reasoning is straightforward:

I. Sally believes no contradiction (i.e. is perfectly rational).

II. "Smith[1]" is DJCR with "he" because the latter is anaphorical on the former.

III. "Smith[1]" is DJCR with "Smith[2]" because the most natural reading of (5) is one in which both tokens are instances of the same name.

IV. "he" is NOT DJCR with "Smith[2]" because that would entail that Sally is being ascribed the contradictory belief that x is someone other than x.

Conclusion. Transitivity fails because "Smith[1]" is DJCR with both "he" and "Smith[2]" but the latter two are not DJCR between themselves.

In this section, I will argue that premises IV is false. That means I will argue that every token in (5) is DJCR with each other. In order to do that I will make some brief remarks about the relations between DJCR and attitude ascriptions.

First, let me highlight the usual distinction between *de re* and *de dicto* attitude ascriptions.

People usually have multiple interests when they ascribe a propositional attitude to someone else. Sometimes they aim to accurately reflect that other person's perspective on the world. When one has that type of objective in mind, one must (minimally) represent the ascribed attitude in a form that would be recognizable by the ascribed subject. However, there are occasions where we are not really concerned with accurately reflecting one's first-personal perspective. Let me give examples of that.

When an attitude ascription is supposed to be interpreted as reflecting the ascribed individual's own perspective, let us call it "de dicto". This is the familiar sense in which Lois Lane does not believe that Clark Kent can fly (although she believes that Superman does). The usual criterion for identifying a *de dicto* ascription is by noticing that it does not allow substitution *salva veritate* of coreferential tokens. Notice how "Lois Lane believes Superman can fly" strikes our understanding as true whereas "Lois Lane believes Clark Kent can fly" does not.

When an attitude ascription is supposed to be interpreted as providing a coarse characterization of an individual's state of mind (not necessarily reflecting his own perspective), we will call it "de re". This is the sense in which Jor-El's utterance of "Lois Lane believes my son can fly" strikes us as true even though Lois herself would deny having any beliefs about Jor-El (she has never heard about him) or his offspring. Since this ascription does not aim at accurately characterizing Lois' perspective on the world, its truth becomes easily admissible. When an ascription is interpreted as *de re*, its truth-conditions remain unchanged under substitution of coreferential tokens, that is, one could truthfully say that Lois believes Clark Kent/Jor El's son/the Man of Steel can fly, given that one has in mind that this ascription aims merely at providing a coarse individuation of Lois' attitudes[24].

The usual philosophical way to flesh out these intuitions is by claiming that the belief relation is systematically ambiguous between two different meanings: *de re* belief and *de dicto* belief. Crucially for our present purposes, one can truthfully ascribe a contradictory belief to a perfectly rational thinker given that this ascription is to be read as *de re*. Since a *de re* ascription presents one's attitudes in a way that is usually not transparent to the ascribed individual, one may believe *de re* a contradictory thought without failure of rationality. Unlike beliefs *de dicto*, beliefs *de re* do not determine whether an individual is rationally faulty. The point here is parallel to one I have made earlier about how *intensional content* is useful for many purposes, but not to individuate cognitive value.

There is an obvious way to connect these previous remarks with the concept of DJCR.

---

[24]Imagine your infant child comes back from school crying and you utter, "Sonny did not like his new Mexican teacher". That utterance could be true (*de re*) even if Sonny does not even have the concept of a Mexican.

Unlike the truth[25] of a *de re* ascription, the truth of a *de dicto* ascription depends on there being DJCR between the ascription's token and some token (pertaining to) of the ascribed individual. More formally:

- A *de re* ascription of a singular belief *a is F* to a subject *S* is true if and only if *S* has a certain belief *b is F* and *a = b*.
- A *de dicto* ascription of a singular belief *a is F* to a subject *S* is true if and only if *S* has a certain belief *b is F*, *a = b* and *a* is DJCR with *b*.

The truth of a *de dicto* ascription, under the supplied definition, requires DJCR between ascribed belief and some correspondent *ascribee*'s belief. Since DJCR makes coreference between its *relata* indubitable, its instantiation explains why *de dicto* ascriptions successfully reflect an individual's own perspective on the state of the world.

Now, let us focus again on (5). Should it be read as a *de re* or *de dicto* ascription?

Let us suppose that it is a *de dicto* ascription. Then, it entails that, according to Sally's own perspective of the world, Smith is someone other than Smith (since "he" anaphorically refers to Smith). However, since no one may rationally believe *de dicto* a contradiction, this would entail that Sally is irrational. Since we are assuming that Sally is perfectly rational, the *de dicto* reading of (5) cannot be true.

Let us now suppose that it is a *de re* ascription. Then, it entails that Sally has some belief that is coreferential with the belief that Smith is someone other than Smith. Perhaps she has the (perfectly rational) belief that the man in the costume is someone other than Smith. More importantly, if (5) is a *de re* ascription, it does not aim to reflect Sally's own perspective on the world. That means that there is no impediment to there being DJCR between "he" and "Smith²", since this instantiation of DJCR in the ascription need not be replicated in Sally's proper belief (as she would express it). In summary: one can *de re* ascribe the belief that x is distinct from x to some subject S without committing oneself to the claim that S is irrational (i.e. believes a contradiction).

I conclude that premise IV is false. It does not follow from "he" being DJCR with "Smith²" that Sally would come out as irrational. If that is the case, we can firmly claim that every token in (5) is DJCR with each other. Since there is no failure of transitivity being exhibited by this representation, there is no argument against (Intrinsicalism) or a Fregean Sense theory.

---

[25]Perhaps I should be talking about the *correctness* of ascriptions, instead of its truth. Nothing of great significance hinges on that and the argument could simply be re-stated in terms of correctness-conditions.

Before finishing this chapter, I will assess a potential criticism that some enemy of Senses could advance regardless of the conclusions in this section (and show why it is misguided).

3.2.1 INDIRECT REFERENCE

Frege (1948) notoriously claimed that words shift their reference when inside attitudinal contexts (which he called "indirect discourse"):

> When words are used in the customary manner then what is talked about is their reference. But it may happen that one wants to talk about the words themselves or about their senses. The first occurs when one quotes someone else's words in direct (ordinary) discourse. [...] In writing we make use of quotes enclosing the word-icons. [...] In indirect discourse we speak of the sense, e.g., of the words of someone else. From this it becomes clear that also in indirect discourse words do not have their customary reference; they here name what would ordinarily be their sense.

If we take Frege at his word, he is literally suggesting that the referent of a token should be given by a disjunctive clause, that is, in customary (non-attitudinal) contexts, it refers to an object; in attitudinal contexts, to some Sense.

Implementing that idea would lead us to the violation of, what Davidson (1968) famously called, *semantic innocence*: the thesis that tokens possess univocal referents in every representational environment in which they can syntactically occur. Take (5), for example. In that representation, we have two instances of the name "Smith", one, outside attitudinal contexts, the other, inside. Here, the denial of *semantic innocence* in accordance with Frege's quotation would entail that "Smith[1]" refers to an object (presumably Smith himself) whereas "Smith[2]" refers to some Sense (of Smith). Thus, two tokens of the same name would fail to corefer. I hope that sounds as awkward to the reader as it does to me.

Even more pressingly, notice how the violation of *semantic innocence* makes it hard to accommodate our intuitions about anaphora across attitudinal contexts. Since "he", in (5), occurs inside the scope of the "thinks that" operator, its reference should be to a certain Sense; its antecedent "Smith[1]", however, occurs in a non-attitudinal context, so it refers to a person. Thus, the Fregean denial of *semantic innocence* lead us to conclude that an anaphorical pronoun may refer to something distinct than its antecedent. A representation, such as "John is honest but Mary believes he is a liar", would come out as true (under that Fregean interpretation) if and only if John is honest and Mary has a belief composed by some Sense (of John) and the property of being a liar. However, which precise Sense of

John should constitute Mary's belief, given that the representation itself is completely silent about any particular mode of presentation?

To be sure, Kaplan (1968) has famously argued that the answer to the last question posed should be "any Sense of John". In other words, "John is honest but Mary believes he is a liar" would be true if and only John is honest and Mary has a belief composed by *any* Sense (of John) and the property of being a liar. With this idea in mind, Kaplan showed how a Fregean could accommodate the intuition that some of our attitudes are *de re*, that is, independent of a particular mode in which their referents are presented. In the context of our present discussion, this Kaplanian idea proves that one could adopt Frege's thesis of indirect reference, violate *semantic innocence*, and still get the truth-condition of representations such as (5) right.

Technically, this is all very satisfying. Nevertheless, notice this troubling consequence: if Frege's thesis is adopted at its word, then tokens which we naturally assume to be instances of a same type will be completely distinct semantically, i.e. distinct in their referents and distinct in their Senses. "Smith[1]" and "Smith[2]", for example, possess no semantic content in common with each other. But, then, in what grounds would the Fregean be able to claim that they are two instances of one and the same name? The Fregean has to say something about this unless he is ready to claim that "Smith[1]" and "Smith[2]" are, contrary to first appearances, instances of distinct syntactic types, that is, that they are instances of distinct names.

I foresee one obvious solution to this problem although, as I shall argue, I do not think it is satisfactory. The Fregean could claim that "Smith[1]" and "Smith[2]" have one semantic content in common with each other. Since "Smith[1]" and "Smith[2]" are both associated with the clause "in non-attitudinal contexts, refer to Smith; in attitudinal contexts, refer to Sense α", there is some semantic similarity between the two tokens which we could use to ground our intuition that they are, in fact, instances of the same name.

The Fregean theory, traditionally construed as a two-factor semantic theory (with referents and Senses), would start to look like a three-factor theory. Tokens not only have referents and Senses but also a third semantic content specifying their referents in distinct representational environments. A relationist philosopher who is already worried about postulating Senses over and above a token's referent will certainly not be happy with this picture. Do we really need all this complication in order to advance a Sense theory that works for attitudinal contexts? Definitely not.

Remember that Frege's motivation for advancing the indirect reference thesis was to explain how two attitude ascriptions could diverge in truth-value when they are constituted by tokens that, in normal contexts, would refer to the same object. That is, the motivation for indirect reference is

the usual fact that singular content is hyper-intensional. However, claiming that tokens shift their referents in different environments is only one way to solve this problem. In fact, this solution to the problem will only seem compelling if one independently holds the following thesis:

(Intensionality) There is never any fact about truth-conditions which is not reflected in *intensional content*/reference.

If one accepts (Intensionality), one has to reduce every difference in truth-conditions between two representations as a difference in their referents. Thus, the indirect reference thesis appears as the natural explanation of that fact. But, why should we stick with (Intensionality)? If one is ready to introduce Senses in one's semantic theory, I see no reason why one should not go ahead and say that some differences in truth-conditions should be explained by a difference in Sense. A natural way to do this would be claiming that some attitudinal contexts are sensitive to the Sense of the expressions in their scope, over and above their referents. The shift that Frege's indirect reference thesis previously effected in a token's reference would then be transposed to the behavior of attitudinal operators.

I think it is important to see that these complexities affect the relationist philosopher as much as they affect the Fregean one. If a relationist philosopher assumes (Intensionality) while also assuming that some attitude ascriptions constituted by coreferential tokens may diverge in truth-value, he will then need to provide a complex justification for that. One could even rephrase Frege's indirect reference thesis in relationist terms. Pinillos, for example, could claim that, inside attitudinal contexts, a token refers to its pattern of P-Linking across the relevant language or idiolect. Since these patterns will be distinct for any two tokens that are not P-Linked with each other, the relationist would achieve the vindication of (Intensionality), but at an exorbitant price. As opposed to that, he could simply claim that attitudinal contexts are sensitive to the patterns of P-Linking a token possesses over and above its reference.

APPENDIX: LEVELS OF IDENTITY/DISTINCTNESS BETWEEN IDENTITY REPRESENTATIONS

Take any two identity representations ⌜ a=b ⌝ and ⌜ c=d ⌝[26]. There are four different levels in which

---

[26] These special quotation symbols are to be read as denoting quasi-quotational contexts, i.e. a variable inside those contexts should be read as a mention of the word it stands for. In other words, "⌜ a=b ⌝" means a mention (not a use) of any identity sentence, so it can be used to make a general point about identity sentences. Analogously, "⌜ a ⌝" is a mention of any

they might be said to be identical or distinct:

De re **Identity**: two identity representations ⌜a=b⌝ and ⌜c=d⌝ are identical *de re* **if and only if** their tokens are coreferential, e.g. ⌜a = c⌝ **and** ⌜b = d⌝ **or** ⌜a=d⌝ **and** ⌜b = c⌝.

Logical Form Identity: two identity representations ⌜a=b⌝ and ⌜c=d⌝ are logical form identical **if and only if either** ⌜a⌝ is DJCR with ⌜b⌝ **and** ⌜c⌝ is DJCR with ⌜d⌝ **or** ⌜a⌝ is INDEPENDENT from ⌜b⌝ **and** ⌜c⌝ is INDEPENDENT from ⌜d⌝.

Weakly De Dicto **Identity**: two identity representations ⌜a=b⌝ and ⌜c=d⌝ are *weakly de dicto* identical **if and only if** ⌜a=b⌝ **and** ⌜c=d⌝ are *de re* identical **and** logical form identical.

Strongly De Dicto **Identity**: two identity representations ⌜a=b⌝ and ⌜c=d⌝ are *strongly de dicto* identical **if and only if** ⌜a=b⌝ **and** ⌜c=d⌝ are *weakly de dicto* identical **and either** ⌜a⌝ is DJCR with ⌜c⌝ **and** ⌜b⌝ is DJCR with ⌜d⌝ **or** ⌜a⌝ is DJCR with ⌜d⌝ **and** ⌜b⌝ is DJCR with ⌜c⌝.

The first relation introduced, *de re* identity, is sensitive only to the referents (intensional content) of the tokens in the representations. In this sense, "Clark Kent = Superman" is *de re* identical to "The Man of Steel = The Man of Steel".

The second relation introduced, logical form identity, is sensitive only to the instantiation (or lack thereby) of DJCR between the tokens in each representation (but not across representations). In this sense, "Clark Kent = Superman" is logical form identical to "Hesperus = Phosphorus". Similarly, "The Man of Steel = The Man of Steel" is logical form identical to "Venus = Venus". In Chapter 2 we defended the thesis that talk of an identity representation being of the form $a = a$ or $a = b$ depends on there being (or not) DJCR between its tokens. With this sense of 'logical' form in mind: all identity representations with the form $a = a$ will be logical form identical (*idem* for identity representations with the form $a = b$).

The third relation introduced, *weakly de dicto* identity (henceforth wdd-identity), is sensitive to intensional content and to 'logical' form (in the sense we have defined). In this sense, "Clark Kent = Clark Kent" is wdd-identical to "Superman = Superman". Similarly, "Clark Kent =

---

referential expression, as opposed to " 'a' ", which is a mention of a particular referential expression, the variable "a".

Superman" is wdd-identical to "Man of Steel = Jor El's Son"[27].

The fourth relation introduced, *strongly de dicto* identity (henceforth sdd-identity), is sensitive to wdd-identity and to the instantiation of DJCR across the two identity representations in question. It is the most demanding sense in which two identity representations can be identical. In this sense, "Clark Kent = Superman" is sdd-identical to "Clark Kent = Superman".

Each of these relations can be given an intuitive exemplification in terms of agreement between believers.

Some subject A who believes that the Man of Steel is identical to itself and some subject B who believes that Superman is Clark Kent agree with each other in the sense that their beliefs are truth-conditionally equivalent (the same state of the world verifies both beliefs). This sense of agreement is captured by the relation of *de re* identity.

Some subject A who believes that Superman is Clark Kent and some subject B who believes that Hesperus is Phosphorus agree with each other in the sense that both hold informative identity beliefs about something. Similarly, some subject A who believes that Clark Kent is identical to itself and some subject B who believes that Hesperus is identical to itself agree with each other in the sense that both hold trivial identity beliefs about something. This sense of agreement is captured by the relation of logical formal identity.

Some subject A who believes that Superman is Clark Kent and some subject B who believes that Jor El's Son is the Man of Steel agree with each other in the sense that both hold informative beliefs about the same thing. Similarly, some subject A who believes that Clark Kent is identical to itself and some subject B who believes that Superman is identical to itself agree with each other in the sense that both hold trivial identity beliefs about the same thing. This sense of agreement is captured by the relation of wdd-identity.

Finally, some subject A who believes that Superman is Clark Kent and some subject B who also believes that Superman is Clark Kent agree with each other in that they both hold the same identity belief. This stronger sense of agreement is captured by the relation of sdd-identity.

A relationist such as Fine (2007) or Pinillos (2011) has a straightforward way to explain these relations in their own favored theoretical vocabulary. All they need to do is substitute any mention of DJCR by the semantic relation their theory postulates, e.g. Fine's "semantic coordination" or Pinillos' "P-Linking". A Fregean also has the tools to provide a correct characterization of these relations in terms of Senses. All she would need to do is substitute any mention of DJCR between two

---

[27] I am assuming "The Man of Steel" and "Jor El's Son" are, analogously to "Superman", directly referential names of Clark Kent.

tokens by *sameness of Sense* between these tokens.

4. INDEXICAL CONTENT IN DIACHRONIC SCENARIOS

As Recanati (2016, p. 1) notes, "there was a time when people took indexicality to be a property of language, and of language exclusively". The idea was that indexicality, similarly to ambiguity, is a property of vehicles of representation, but not of the represented contents themselves. One way in which this assumption manifested itself was by the dismissal of the thesis that indexical expressions have conceptual analogues, or that some thoughts are themselves indexical.

The orthodoxy on those matters has changed a lot since the publication of important works by Prior (1959), Castañeda (1999) and Perry (1993). These authors independently pointed to a curious fact about indexical representations: their cognitive value seems essentially connected to the indexical tokens that constitute them. Their argument is straightforward: one can believe whatever is expressed by a certain indexical representation without believing whatever is expressed by a non-indexical cointensional representation. Thus, Brigitte, an amnesiac at the hospital, can sincerely utter, "I am an amnesiac", without being sure whether she should assent to, "Brigitte is an amnesiac"[28]. Even though the two singular representations are cointensional, their cognitive value is distinct. In other words, "I" and "Brigitte" are not DJCR in Brigitte's mouth, since she would be rational in wondering whether they corefer. The same argument applies to cointensional representations which differ merely by the substitution of an indexical expression for another. Thus, one can sincerely utter, "He is making a mess", while pointing to a reflection of himself in a mirror, while being unsure whether to assent to "I am making a mess". Furthermore, which indexical constitutes a representation is essential to its import in psychological explanation and action-guidance. Thus, when I think whatever would be expressed by me uttering "I am being attacked by a bear", I try to defend myself, but when someone else thinks, about me, whatever would be expressed by uttering "He is being attacked by a bear", one runs to get help.

In this chapter, we will analyze the question of indexical content in diachronic scenarios. As it will be shown, it is notoriously hard to individuate the cognitive value of indexical representations when they are made in temporally distant moments.

4.1 THREE DIACHRONIC SCENARIOS

Consider these three scenarios involving Alice, Bob and Carol:

(Alice) Alice sincerely utters "Today is a wonderful day" at Monday before going to sleep. As usual,

---

[28] Conversely, we can conceive of scenarios in which Brigitte knows that *Brigitte is an amnesiac* without knowing that *she herself is an amnesiac*.

she wakes up at Tuesday and, intending to re-express the same belief she had previously expressed by her "Today"-utterance, she utters "Yesterday was a wonderful day".

(**Bob**) Bob sincerely utters "Today is a wonderful day" at Monday before going to sleep. Unbeknownst to him, someone drugs him in his sleep. The drug makes him sleep for two days, thus, he only wakes up at Wednesday. Without realizing that two days had passed, he utters "Yesterday was a wonderful day" intending to re-express the same belief he had previously expressed by his "Today"-utterance.

(**Carol**) Carol sincerely utters "Today is a wonderful day" at Monday. Before going to sleep, she willingly takes a very heavy dose of sleeping pills. She ends up sleeping for two days and only wakes up at Wednesday. Since she remembers having taken so many sleeping pills, she is unsure for how many days she slept. Nonetheless, she utters "Yesterday was a nice day" with the sole purpose of re-expressing the same belief she had previously expressed by her "Today"-utterance.

Each character in these scenarios produce two utterances: a "Today" utterance on Monday, and a "Yesterday" utterance one or two days later. Similarities apart, there are plausible normative differences between Alice and Bob, on one hand, and Carol, on the other.

Alice's case, to begin with, is one where everything goes as well as it could. Her intention is to re-express a "Today" belief when that day is no longer present. It seems intuitive that the main way to achieve that feat is by transforming the verbal expression of that belief in some suitable way. Since only one day had passed, the suitable transformation would be to use the indexical "Yesterday" in place of "Today". This very ordinary intuition was noticed by Frege (1997, p. 10), who remarked that "if someone wants to say the same today as he expressed yesterday using the word 'today', he must replace that word with 'yesterday'. Although the thought is the same its verbal expression must be different in order that the change of sense which would otherwise be effected by the differing times of utterances may be cancelled out".

From a first-personal point of view, Bob's situation is completely analogous to that of Alice. Throughout the whole episode, things seemed to Alice exactly as they seemed to Bob. Regardless of that, he is not as successful as she is in re-expressing his belief, since his "Yesterday" utterance ends up referring to Tuesday, a day in which Bob did nothing but sleep. Still, I think it is undeniable that, insofar as Alice is perfectly rational, Bob should not come out as less rational than she is.

Carol's scenario deviates from the previous two in that she has strong reasons to think that

uttering "Yesterday" after waking up may fail to pick out the day that she intends it to pick. Nonetheless, she goes ahead and utters it. It goes without saying that she fails to pick out Monday by means of "Yesterday" and ends up referring to a day in which she was asleep. Unlike Bob, who committed a similar mistake, Carol seems to be guilty of some kind of normative mistake: she has clearly done something she knew she should not have. That is so because she acted against strong reasons that she consciously possessed (and which Bob completely lacked).

There are many contrasting ways to individuate the cognitive value of these three individuals' indexical tokens. As I have been defending, the identity/difference of cognitive value between two tokens supervenes on the instantiation (or lack thereby) of DJCR. Two tokens being DJCR, for its part, is a matter of a thinker having rational entitlement to believe that they corefer (if they refer at all).

In the next few sections, I want to go over four distinct theories that one might defend regarding the sameness and distinctness of the cognitive contents they have expressed with their two utterances. Each theory will differ from the others in some particular manner. I will go over several reasons why one should not accept the presented theories and leave my own preferred view for the end.

4.2 THE NON-FACTIVE THEORY: BOZICKOVIC (2015)

Two tokens are DJCR when a speaker has an entitlement to the belief that they corefer (if they refer at all). Since it is impossible that temporal indexicals, such as "Today" and "Yesterday", fail to refer, we need not be concerned with the possibility that they come out empty[29]. Thus, two temporal indexical tokens are DJCR with each other if and only if one (who understands them) is rationally constrained to believe that they corefer.

With that in mind, let us turn our focus back at Alice, Bob and Carol. Suppose that Alice, after waking up on Tuesday, decides to re-express the belief she expressed at Monday by her "Today" utterance but is not confident as to whether uttering "Yesterday was a wonderful day" will succeed in achieving that feat. Is her lack of confidence rationally warranted? There is no unanimous answer here, but one thing strikes me as certain: someone who wonders whether one has lost track of time without having any obvious reasons to that effect is, as least, extravagantly over-zealous. Perhaps Alice heard about Bob's story and now wonders whether she herself was drugged during the night (even though there is no reason why Bob's life would have any import on her own). The question, then, is: is epistemic over-zealousness irrational? At least in some cases, it seems to be, e.g. one who always

---

[29]I am obviously disregarding the possibility (?) that one, perhaps God, might utter "Yesterday" shorty after the beginning of Time.

suspected that life is nothing but the dream of a butterfly would universally be considered as a lunatic.

Let us then suppose, for the sake of the argument, that Alice would be irrational to consider being in Bob's scenario. This implies that it would then be irrational for her to consider the possibility that "Today" and "Tomorrow" refer to distinct days. In other words, her indexical tokens would come out as DJCR. The same rationale would apply to Bob's case. Since he has no reasons to suppose that he is a victim of such an absurd scenario (as he in fact is), it would be irrational for him to consider the possibility that his two tokens fail to corefer. Thus, they are DJCR. As for Carol, things are different. Since she has strong reasons to consider the possibility that her two tokens fail to corefer, they come out as independent.

This would be our result: Alice and Bob's tokens are DJCR with each other whereas Carol's are independent. This means that the cognitive value of Alice and Bob's two utterances is identical, whereas Carol's two utterances are cognitively distinct. If one accepts that cognitive value is the semantical notion that correctly individuates mental states, such as thoughts and beliefs, the outcome is that Alice and Bob successfully re-express their "Today" beliefs by means of their "Yesterday" tokens while Carol does not.

As is so common in philosophy, this intuitive theory lead us to unintuitive conclusions. Notice that this theory has the strange consequence that two tokens might be DJCR without them being coreferential in the first place. That is the case for Bob's two indexical tokens. Even though he seems to be rationally entitled to the belief that they corefer (and he would be irrational to think otherwise), they refer to distinct days, i.e. Bob's "Today" token refers to Monday and his "Yesterday" token refers to Tuesday[30]. However, there is surely something awry with the idea that two tokens can be *de jure coreferential* without them being coreferential in the first place.

I will henceforth call this theory "the Non-Factive theory", since it denies the following thesis:

(**Factivity**) If two tokens are DJCR, then they are coreferential (if they refer at all).

A recent example of a Non-Factive theorist is Bozickovic (2015), who argues that the reference of an indexical token should be wholly dissociated from its cognitive value, since the former depends exclusively on its linguistic meaning (and the context on which it was uttered) while the latter depends on facts internal to the utterer, such as whether one is representing a day "*as the same* day

---

[30]I will simply assume that the semantic reference of an indexical supervenes exclusively on its linguistic meaning (its Kaplanian *character*, as in Kaplan (1989a)). I do not see any reasons to refrain from that widely held assumption.

from one occasion to the next, […] thinking of it via the same sense, under the same mode of presentation" (ibid., p. 479-80). Bozickovic's reasoning seems to be based on the fact that we can always be mistaken about the context of utterance within which we are presently situated (just as Bob is mistaken about his present date) and, because of that, the continuity of our cognitive life – which is something internal to a thinker - should not depend on us having true beliefs about contingent states of affair. As a result, two tokens may be cognitively identical (due to facts cognitively accessible by its producer) while being referentially distinct (due to facts external to its producer). Thus, even though Bob mistakenly believes it is presently Tuesday, when it is Wednesday, his "Yesterday" token is cognitively identical to his "Today" token, since, Bozickovic claims, he is using them to represent a day *as the same* from one occasion to the other.

While I think Bozickovic is right in trying to make cognitive value come out independent of epistemically opaque issues, i.e. external facts that are not reflectively accessible such as whether one has kept track of time, there are two deep problems with his theory.

Firstly, it counts someone as re-expressing a belief when one represents it, at a later time, *as the same*. Since Bozickovic presents no explanation of what representing *as the same* means, and how we should assess whether one is doing that, I do not see any way to understand his proposal except by interpreting "representing *as the same*" as representing with DJCR[31]. It should be obvious that his proposal would then be guilty of circularity, since our initial predicament is precisely to establish whether Alice, Bob and Carol's tokens are cognitively identical or not.

Secondly, even if Bozickovic had a clear criterion for "representing *as the same*" which enabled us to assess tokens for sameness of cognitive value independently of their being DJCR with each other, he would still have to deal with the fact that his notion of cognitive value is not factive. To see how deeply problematic this is, notice that the main *raison d'être* of cognitive value, i.e. a semantic value over and above referential value, is to individuate one's thought contents in a way that enables the assessment of one's rationality and which provides reasons for one's actions. For example, when one desires to meet up with John and believes that going to the park is an effective way of meeting up with John - where both instances of "John" are cognitively identical - one has a reason to go to the park. Furthermore, it would, *ceteris paribus*, be irrational for that thinker not to go there. The correctness of this assessment of a thinker's rationality presupposes that the cognitive identity of the two "P" tokens guarantee that they will converge on the same individual, since there would be no obvious reason to go to the park if the two tokens of "John" referred to distinct individuals. One could

---

[31]As a matter of fact, "representing as the same" is Fine's (2007) preferred tag for what we are calling "*de jure* coreference".

even claim that it would be rational for that thinker not to go to the park, since he has no guarantee that acting thus would help him achieve his goals. In the absence of a more complicated theory, the rejection of (Factivity) deprives cognitive value of the normative import that it is supposed to have.

Bozickovic takes his Non-Factive theory to be the best way to ensure that thinkers who lose track of time, such as Bob, are nonetheless able to retain the indexical beliefs they had before becoming confused. His argument is that, once one enforces (Factivity), one must conclude that Bob is merely under the illusion of retaining the belief he expressed at Monday by his "Today" utterance. Notoriously, Evans (1981) was perfectly aware of that predicament but saw no problem in acknowledging it. Let us take a look at Evans' theory before summing up the state of the discussion.

4.3 THE NON-TRANSPARENT THEORY: EVANS (1981)

According to (Factivity), no tokens can be cognitively identical without them also being coreferential. One particular consequence of it is that Bob's tokens come out cognitively distinct. Now, add to that the plausible thesis that Bob sincerely takes himself to be re-expressing his "Today" belief by means of his "Yesterday" utterance. From these two theses, it follows that Bob is mistaken about the sameness of the cognitive value of his own thoughts, i.e. they are cognitively distinct but he believes they are not. In other words, the adoption of (Factivity) together with some other independently plausible claims entails that Bob is merely under the illusion of having retained his "Today" belief when, in reality, he lost that belief as soon as he lost track of time.

This description of Bob's thoughts makes belief retention turn out non-transparent, i.e. epistemically opaque, since assessing whether one has retained a belief presupposes assessing whether one has knowledge about the context in which one is located. Since one can always be mistaken on *a posteriori* grounds about the context in which one is located, even the most rational subject can be under the illusion of having retained a belief when she has not.

However, it seems very fishy to claim that Bob has lost the belief he expressed at Monday by his "Today" utterance, after all, there is no moment in his story in which he changes his mind about that day (which he remembers) being a wonderful day. Surely Bob still vividly remembers that day and still thinks it was a wonderful one, even though he is incorrect about its temporal location.

Regardless of admitting all those counter-intuitive conclusions of a Non-Transparent theory, Evans (1981, p. 311, ff. 21) had did not hesitate to endorse it:

> I see no more strangeness in the idea that a man who loses track of time cannot retain beliefs than in the idea that a man who loses track of an object cannot retain the beliefs

about it with which he began. If one has in fact lost track of time without knowing it, then one could think that one had retained one's beliefs when one has not.

Evans' reasoning is based on two points which I will claim to be mistaken. First, he notices that we cannot think of a day as it being Today when that day is no longer present, that is, he notices one cannot re-express at Tuesday a "Today" belief held at Monday by simply making use of the same representation twice over. While this observation is perfectly correct, Evans derives from it the unwarranted conclusion that the only way to re-express, and thus, to retain, at Tuesday a "Today" belief held at Monday is by making the proper modification in the indexical representation one uses to express that belief, that is, to think of that day by means of "Yesterday". Now, if the only way to retain a "Today" belief one day later is to convert it into a "Yesterday" belief, people like Bob would end up having lost the indexical beliefs they started with. I claim that Evans' conclusion is unwarranted since there is no entailment from

(I) One does not retain the belief expressed by the representation "Today is a wonderful day" at Monday by tokening "Today is a wonderful day" at Tuesday.

To

(II) The only way to retain the belief expressed by the representation "Today is a wonderful day" at Monday is by tokening "Yesterday was a wonderful day" at Tuesday.

To be sure, (I) is compatible with there being multiple ways to retain a "Today" belief, while (II) entails that there is only one privileged way of doing that. In later sections I will claim that we can retain our indexical beliefs by means of converting them into memory beliefs (which is completely different from converting it into a "Yesterday" belief), but that will have to wait a while longer.

The second point which I take to exhibit a failure in Evans' reasoning is his inference from the importance of our tracking abilities to the claim that these tracking abilities are a precondition of belief retention. I take the premise of this inference to be right on the mark, that is, I agree with Evans (ibid., p. 309) that "no one can be ascribed at t a belief with the content 'It is now PSI', for example, who does not have the propensity, as time goes on, to form beliefs with the content 'It was PSI just a moment ago', 'It was PSI this morning', 'It was PSI yesterday morning' etc." and "that a capacity to

keep track of the passage of time is not an optional addition to, but a precondition of, temporal thought". Where I disagree with Evans is about whether these facts entail that belief retention – or, more generally, the identity conditions of indexical beliefs across time – supervenes on correctly keeping track of the context. Whereas Evans thinks it does, I claim that it does not. The thesis that tracking abilities are a cognitive precondition for having temporal thoughts only entails that someone who "systematically loses track of time and never knows his own position in time" is not capable of properly tokening any temporally indexical representations (Branquinho, 2000 p. 6). However, it does not entail that the cognitive value of indexical tokens supervenes on that cognitive capacity being correctly executed nor that someone who occasionally loses track of time will thus be under the illusion of having retained her beliefs. It might very well be the case, as far as the premise of this inference is concerned, that someone who only rarely loses track of time, such as Bob, is perfectly able to retain his temporal beliefs and Evans has failed to provide any reasons to doubt that this is correct.

Now, even if these failures in Evans' argumentation could be amended, there is a deeper worry which threatens the explanatory potential of his theory. Since Evans claims that thinkers can be equivocated about the cognitive value of their own thoughts, he ends up rejecting the principle of Transparency in its most general form:

(Transparency) If two of a thinker's tokens are DJCR, then that thinker must be able to know that they do without recourse to experience; if two of a thinker's tokens are independent, then that thinker must be able to know that they do without recourse to experience.[32]

This principle implies that every thinker can know, immediately and without recourse to experience, whether his or her inferences are valid or not. Since to neglect immediate knowledge is a symptom of irrationality, this principle implies that no thinker can be fully rational while committing logical mistakes in reasoning.

To deny (Transparency) entails that it is possible to produce a textbook case of a fallacious argument even while in maximum control of one's rationality. Under the denial of (Transparency), assessing arguments for their soundness and for their validity are similar in that both require the agent

---

[32]I have intentionally mimicked Boghossian's (1994, p. 36) influential exposition of the Transparency principle, although he put it in terms of thoughts "possessing the same content" or "distinct contents" – which, for all relevant purposes, I see as meaning the same as two tokens instantiating DJCR:

> (a) If two of a thinker's token thoughts possess the same content, then the thinker must be able to know *a priori* that they do; and (b) If two of a thinker's token thoughts possess distinct contents, then the thinker must be able to know a priori that they do.

to go check the external world. This predicament is considerably worrying since we typically take the logical properties of an argument to be graspable by its form, while grasping its form does not depend on any empirical information:

> If we could be wrong, on empirical grounds, about the contents of our own thoughts, then we could be wrong, on empirical grounds, about the validity of our reasoning, and this seems incompatible with the idea that we can separate the assessment of reasoning from assessment of the truth of the premises on which the reasoning is based. (Stalnaker, 2008, p. 114-5)

Since the instantiation (or lack thereby) of DJCR would cease to be determinable without recourse to experience, cognitive value would become an empirical matter. Furthermore, because our knowledge of empirical facts is fallible by nature, even the most rational of thinkers could be guilty of committing trivial logical mistakes. Rationality and validity would not converge.

Non-Transparent theories seem to betray their own objectives, since they neutralize one of the main motivations for acknowledging semantical aspects of singular representations that are more fine-grained than their intensional contents. That is, one of the reasons why intensional content was claimed to be insufficient for many theoretical purposes was that it is not transparent, i.e. a rational subject can be oblivious to the fact that two representations are cointensional. Recanati (2012, p. 117), expressing this precise point, writes:

> If modes of presentation [cognitive values] are not themselves transparent, there is no reason to move from pure referential talk to mode of presentation talk in the explanation of rational behavior. The appeal to senses (as opposed to sheer reference) in psychological explanation presupposes the transparency of senses as opposed to the non-transparency of reference.

In defense of Evans, some could complain that this diagnostic is an over-reaction. "Surely", they would say, "there are ways to reject (Transparency) without creating anarchy in rationality assessments". To be sure, I think this is perfectly true. Indeed, most philosophers who reject (Transparency) go on to argue that cases of *validity mistakes concurrent with maximum rationality* are special exceptions, not the norm[33].

Be that as it may, I am not advancing a transcendental demonstration of (Transparency). It is enough for my purposes to argue that a Non-Transparent theory creates a gap between cognitive value and normative import, since one cannot straightforwardly use an (epistemically) opaque notion to

---

[33]In chapter 2, we have briefly seen two examples of philosophers who claim something in these lines, Gerken (2013) and Faria (2009).

assess a thinker's rationality or to rationalize one's actions. To illustrate, remember the thinker who desires to have a meeting with John and believes that going to the park is an effective method of having a meeting with John. Even though we want to say that the most rational thing for him to do is to go to the park, Evans' theory – more generally, any Non-Transparent theory - entails that this thinker could be mistaken about his desire being about the same subject matter as his belief, so that there is no guarantee that going to the park will help her achieve her objective. Since it is possible that this thinker is equivocating between different singular concepts, it is *prima facie* also rational for her not to go to the park. Thus, cognitive value and normative import seem to go separate ways.

The astute reader has surely noticed by now that the consequences of rejecting (Transparency) are identical to those of rejecting (Factivity). Both strategies dissociate cognitive value from the reason-giving role it originally was supposed to have, and they do it for similar reasons. On one hand, a notion of cognitive value which is Non-Transparent, is one which we can trust but have is not easily accessible, on the other, a notion of cognitive value which is Non-Factive is one we have easy access to but which we cannot trust. (Factivity) and (Transparency) are theoretically complementary. A notion of cognitive value that respects one, but not the other, will necessarily be impaired in a significant way. It will be either a semantic value that may elude even the most rational of thinkers, or a value that may not converge on truth. In conclusion, these two principles should be accepted in conjunction or be gotten rid of altogether.

4.4 THE W-DJCR THEORY: RECANATI (2012, 2013a, forthcoming)

In a series of publications, Recanati[34] has argued for a theory according to which Mental Files play most of the roles that we have so far associated with cognitive value. Mental Files are mental particulars by means of which thinkers have singular thoughts and store singular information. They are the mental analogues of referential expressions, such as proper names, indexicals, and demonstratives.

Every Mental File is based upon at least one "acquaintance relation", or, as Recanati prefers to call them, "epistemically-rewarding relations" (henceforth ER-relations), i.e. "a relation such that, when one stands in that relation to some object, one can gain information from the object through the relation" (Recanati, 2013, p. 3). These relations determine the identity conditions of Mental Files, that is, two Files are identical if and only if they exploit the same ER-relation; otherwise, they are distinct. Furthermore, ER-relations also determine the referent of the Mental Files with which they are

---

[34] His 2012 book *Mental Files* contains the more important arguments in favor of his picture. In his 2013b and 2015b provides enlightening replies to a few of his critics. In this section, we will mostly be concerned with his 2013a and forthcoming which focus on the question of cognitive dynamics. His 2016 extends the discussion to communicative scenarios.

associated: a File *A* tokened by a thinker refers to an object *x* if and only if *x* is related to that thinker by the ER-relation constituting the File *A*. Since the identity conditions of Mental Files supervene on their reference-fixing ER-relations, it is impossible that two tokenings of the same Mental File fail to corefer (if they refer at all). Thus, when two tokens are associated to the same Mental File, they are DJCR (cognitively identical).

The ER-relations constituting our Mental Files, specially those Files associated to indexical and demonstrative expressions, are analogous to the linguistic meaning (*character*) of context-dependent expressions. Take the indexical "I", for example. Its character is the rule that every token of "I" refers to whomever produced it. Similarly, the linguistic meaning of "Today" is the rule saying that every token of "Today" refers to the day on which it was produced. Characters are associated to indexicals at the level of expression-types, i.e. it is something every token of the same indexical has in common. However, indexicals only have a referent at the level of expression-tokens. To determine the referent of an indexical token, such as Alice's "Today" token, one puts its character together with the context on which it was produced in order to reach a reference-fixing rule along the following lines: Alice's token of "Today" refers to *x* if and only if *x* is the day on which Alice's token was produced. These reference-fixing rules are usually called token-reflexive since, according to them, the referent of a token must possess a relational property to the token itself.

Even though the linguistic notion of a character does not apply to Mental Files, their associated ER-relations function in a very similar manner. Indexical Files, i.e. the Files associated with indexical expressions, are type-identical according to the type of ER-relation to which they are associated. The SELF file, for example, is that type of File by means of which thinkers acquire information about oneself in that special manner in which only we ourselves are able to acquire information (about ourselves), e.g. proprioception, introspection etc. Every thinker possesses a SELF file, although no SELF file is the same across different thinkers. They are type-identical, because they exploit the same type of ER-relations, but token-distinct, because each of them exploits a particular ER-relation directed at a particular individual. Similarly, there are Temporally Indexical Files, such as the NOW file or the TODAY file. The former exploits a type of ER-relation which individuals have a to a time when (and only when) that time is present, whereas the latter exploits a type of ER-relation which individuals have to a day when (and only when) that day is present. Similarly to indexical expressions, Indexical Files only have a referent at the level of their tokens, i.e. to determine the referent of a NOW token, one "adds" its type of ER-relation to the context on which it was uttered to get, as a result, a token-reflexive rule along the following lines: a token of NOW refers to *x* if and only if *x* is the time of tokening of this NOW-token.

The ER-relations that individuate Indexical Files are very ephemeral and unstable. As soon as one day ceases to be present, the ER-relation which enabled thinking about it by means of the TODAY file ceases to exist. That is, the exercise of the TODAY file in these conditions would end up referring to a distinct day. Since the identity of Mental Files is factive (in regards to coreference), the exercise of the TODAY file in these conditions would be the exercise of a distinct file altogether.

As some of authors have noticed, this fine-grained individuation of Mental Files makes it very hard to explain how thinkers can retain their indexical attitudes when the ER-relations grounding these attitudes have ceased to exist (Onofri 2014, Ninan 2014). To overcome that problem, Recanati (2012, p. 81) appeals to the process of *conversion* with which the "information stored in a file is transferred into a successor file when the ER-relation which sustains the initial file comes to an end". *Conversion* is at place, for example, "when a demonstrative file gives way to a memory demonstrative" (ibid.). This particular instance of *conversion*, i.e. the conversion of a demonstrative file into a memory file, explains how we are able to retain demonstrative beliefs about an object when we are no longer in a demonstrative relation to these objects:

> At *t*, the subject sees the object, and can store information derived from the perceptual episode; […]. As the episode comes to an end, the subject stays, through memory, in contact with the object, but the relation to the object is different. Since the relation changes, I said that the perceptual file is replaced by a memory file, but the word 'replacement' hides the continuity between the memory file and the initial perceptual file: in a certain sense, it is the same file—the same body of information—that changes its status as the ER relation on which it is based changes. It is that continuity which the notion of *conversion* highlights.

In more recent papers, Recanati (2013a, forthcoming) notices that *conversion* is only one of multiple *dynamic operations* which thinkers execute on their Mental Files - specially those files based on unstable ER-relations, since those require a lot of updating on the part of the subject in order to retain their attitudes across different cognitive states. Recanati (forthcoming, p. 17) distinguishes five types of dynamics operations on files: (i) conversion of one file into another, (ii) incremental conversion, (iii) fusion, (iv) fission, and (v) deployments of the same file at different times. Let us briefly go over each one of them.

Conversion, as we have already seen, is that process according to which one file is transformed into another which inherits its content. Recanati discusses a couple of paradigmatic instances of that process: the *conversion* of demonstrative files into memory files and the *conversion* of NOW files into THEN files. These cases are composed by two distinct moments. A first moment where a subject acquires information through a unstable ER-relation, i.e. being an object

demonstratively/perceptually related to a subject & being a time presently related to a subject – and the subsequent transference, when the first relation cases to hold, of that information to a new file which is based on a more stable memory-based ER-relation. In *conversion* processes, memory seems to be doing most of the hard work: a memory file supersedes an unstable file when the latter's ER-relation ceases to hold. On one side, it can be claimed that the memory file is distinct from the file it supersedes – since their constitutive ER-relations are distinct. On the other side, it can be claimed that the memory file just is the same file as that which it superseded – since they are related by means of a *dynamic operation* and this is the only way to retain a belief after its constitutive ER-relation has ceased to exist.

I find it remarking that, in one passage (forthcoming, p. 17), Recanati seems to assume (without arguing for) that transforming a TODAY file into a YESTERDAY file is as much an instance of *conversion* as the other two cases aforementioned. I find this particularly remarking because it is not obvious whether the purported *today/yesterday conversion* has as much to do with memory as the demonstrative and *now/then* cases. We will have more to say about this later on.

The second *dynamic operation* postulated by Recanati, *incremental conversion*, is that process by means of which a file acquires distinct ER-relations over and above the one(s) it originally had. A paradigmatic instance of that process is our ability to integrate information cross-modally, i.e. through different sensory modalities. Imagine learning, in the same perceptual episode, that an object is sturdy by touching it and that it is translucent by seeing it. In those very familiar cases, we compute information coming from two (or more) sensory modalities as being primitively integrated, that is, in no point we judge *the object touched = the object seen* before making inferences that treat the tactile demonstrative and the visual demonstrative as being the same concept. Recanati's explanation is that, in those cases, a simpler file grows an additional ER-relation and becomes a composite file. In the cross-modal case we have introduced, Recanati argues that the perceptual file employed is, in a first moment, constituted by the tactile ER-relation, and in a second, by the tactile and the visual ER-relations. *Incremental conversion* takes a simple file constituted by a certain ER-relation α and outputs a composite file constituted by multiple ER-relations α, β etc. The structure of the simpler file and the post-increment files reflect the structure of tokens that are *weakly de jure* coreferential to each other. Remember that two tokens are W-DJCR when it is rationally impossible that they have distinct referents. The Fregean explanation for that fact was that two tokens are W-DJCR when their Senses have a part in common. Since one can see Recanati's ER-relations as playing the explanatory role of Senses (as far as our present concerns go), it is easy to conclude that two files are W-DJCR when their constitutive ER-relations have a part in common. Thus, *incremental conversion* is a dynamic operation which creates a composite file from a simpler file, whereas these two files will come out as W-DJCR to

each other (but not DJCR).

Fusion and *fission* are similar to each other and to *incremental conversion*. Imagine that one starts wondering whether one's tact is acquiring information about the same object that one is seeing (in one's own hands). Since rational doubt entails cognitive difference, one would have in that case two distinct files, a tactile file and a visual file. These files have resulted from a *fission* of a composite file. *Fission,* thus, is that process by means of which we separate the ER-relations constituting a composite file and create multiple simpler files. *Fusion* is the inverse process, i.e. when one integrates two distinct files into a larger composite one. The only difference between *fusion* and *incremental conversion* is that the former unites two previously existing files, while the latter increments a file with a new ER-relation that was not being previously exploited. It follows that these processes, similarly to *incremental conversion*, only ensure W-DJCR between its input and output files.

Finally, Recanati takes, as a paradigmatic example of *deploying the same file at different times*, the case of someone who, at t^1, acquires visual information about an object. At t^2, the object disappears but the subject hallucinates that it is still there. At this later time, his demonstrative file will seemingly fail to refer, because it presupposes that (i) there is some object in the demonstrative relation to the subject and that (ii) this object is identical to the one at t^1. Since these presuppositions are doubly violated, it seems plausible to claim that the deployment of that demonstrative file at t^2 fails to corefer with its previous deployment at t^1. As Recanati goes on to conclude, "two deployments of the same file at different times are not bound to corefer if they refer at all"; they will only be W-DJCR, i.e. bound to corefer if *both* refer.

After making those illuminating distinctions, Recanati notices that all of those *dynamic operations* ensure, at most, W-DJCR between its input and its output files. This is an interesting conclusion, since it enables him to carve a middle way between Bozickovic's (2015) Non-Factive and Evans' (1981) Non-Transparent theories. Unlike Bozickovic, Recanati does not claim that one's indexical tokens in diachronic scenarios, e.g. Alice's "Today" and "Yesterday" tokens, are cognitively identical. This would require DJCR between them, whereas his theory entails that these tokens will be, at most, W-DJCR. On the other hand, Recanati claims, *contra* Evans, that tokens which have been submitted to *dynamic operations* will never refer to distinct objects – even if the subject has lost track of some context change. Since they are W-DJCR, as opposed to independent, their coreference will be guaranteed (as long as *both* tokens refer).

Recanati manages to save (Factivity) since no tokens will be cognitively related while failing to corefer (if both refer at all). He (2012, ch. 10-11) also goes at great lengths to defend (Transparency), and is able to do so because, according to his theory, thinkers have immediate access to

the identity conditions of their own Mental Files. Cases of subjects who lose track of time would not threaten the Transparency of Mental Files since Recanati reinterprets those cases as necessarily involving one file which succesfully refers and its *dynamically operated* counterpart, which fails to refer due to violation of its presuppositions. Since these two files are only W-DJCR, the instantiation of a case 3 - i.e. where one file refers while the other does not – does not mean that a subject is confused about the content of her own thoughts.

I will now make some critical comments directed at Recanati's theory. First, I assess his claim to the effect that *dynamic operations* only ensure W-DJCR between its input and output files. More particularly, I am worried whether we can make sense of there being W-DJCR between files created by means of *conversion* and *redeployment*, since those processes do not seem to output files whose ER-relations have a part in common to the ER-relations of its inputs – what I take to be a constitutive property of W-DJCR. Secondly, I will claim that the process of *redeploying the same file at different times* boils down to the process of *incremental conversion*. Thus, I will claim that *redeployment* is not a basic dynamic operation, although it is reducible to ones that are more basic. Thirdly, we will be left with the predicament of explaining how can *conversion* output files which are W-DJCR to its predecessors even though there does not seem to be any ER-relation which they have in common. Finally, I will argue that *conversion* ensures proper DJCR, *contra* Recanati, and that this is so because of the way information is stored through memory demonstratives. I will make an analogy between memory demonstratives and deferential expressions to motivate that claim. In the end, I will point out that, if my observations are correct, the case of "Today" and "Yesterday" tokens made vivid by Alice, Bob and Carol should not count as an instance of a dynamic operation.

4.4.1 REDEPLOYMENT

Recanati (forthcoming, p. 17) argues that "two deployments of the same file at different times are not bound to corefer if they refer at all; they are only bound to refer to the same thing if they *both* refer (W-DJCR)". This entails that the mere re-tokening of a Mental File is itself a dynamic operation: a cognitive transformation in one's thought contents. In this subsection, I want to clarify how W-DJCR arises from redeployment of files.

Recanati's main case study for that thesis mentions one who is visually attending to some object and does not realize that, at some point, the object disappears or is substituted by a qualitative identical. Suppose that, at t^1 (before the disappearance/substitution), the subject is acquiring information by means of a demonstrative file. At t^2, something weird happens: the object disappears and the subject hallucinates that it is still there or the object is surreptitiously substituted by a

qualitative identical one. The question is: what happens to that demonstrative file at t²? Recanati (ibid., p. 14) claims that it does not refer anymore, since its deployments presuppose two facts: "that there is an object to which the subject is attending, and that whatever the subject is attending to is the same thing she has attended to in previous deployments of the file". If the object was substituted, the second presupposition is violated; if the object disappeared, both presuppositions are violated. In both cases, at least one presupposition is violated and Recanati argues that this makes the demonstrative file come out empty. But does it?

A first source of skepticism about Recanati's story is that these presuppositions, which the confused thinker violates, do not seem to belong to the ER-relations determining the reference of the demonstrative file. As Recanati argues many times (e.g. 2012, p. 60-64), the deployment of a Mental File has to meet certain *normative requirements* if it is to count as a felicitous deployment. For example, one must not open a demonstrative file if one is not demonstratively related to an object or if one is related to more than one object. However, it is not obvious whether these requirements should be construed as part of the reference-fixing conditions of a file. As we have seen, the reference of a file is determined by its ER-relations and these seem, *prima facie*, to be independent of the conditions a rational thinker should respect while exploiting them.

Regardless of that, one could argue that the ER-relation governing demonstrative files themselves are such that they require the existence of one and only one object as the source of information one is acquiring. One way to do that which seems to be textually supported by Recanati's general position is by claiming that the ER-relation governing demonstrative files is representable as

$$\lambda x \lambda y(x \text{ is visually keeping track of } y)$$

where *visually keeping track of something* is a temporally extended relation which intrinsically presupposes that one is visually attending to one, and only one, object throughout a certain perceptual episode. That a dynamic relation such as this grounds our demonstrative thoughts seems to be corroborated by cognitive scientists' research on infant and adult perception, where it is claimed that our cognitive system represents spatiotemporally coherent objects even in its most incipient (and cognitively encapsulated) stages, i.e. that our cognitive system, by nature, keeps track of spatiotemporal objects (Scholl 2001, Pylyshyn 2007).

However, if that story is correct, it is difficult to explain how a thinker could start wondering, in the midst of a perceptual episode, whether her demonstrative file is picking out, at that moment, the same object it was picking out in the beginning. Rational doubt entails cognitive

difference, so that thinker would need to have two distinct files at that point. She could reach these two files by executing a *fission* on her demonstrative file. However, the relation of *visually keeping track of something* is not a composite relation, and I do not see how one could perform *fission* on a file that is based on a simple ER-relation. How could a simple ER-relation be divided into two others? An immediate reply would be that, perhaps that tracking relation is covertly composite:

> *x* is visually keeping track of *y* [at t^2] if and only if *x* is visually attending to *y* [at t^2] and *x* was visually attending to *y* [at each $t < t^2$, for all t contained in the duration of the perceptual episode]

If that is a correct representation of the *visual tracking* relation, it is clear how one could perform *fission* on it and output two distinct files. *Fission* would output one file for the object being visually attended to at t^2 and another for the object which was visually attended to previously to that. I am sympathetic to that construal of demonstrative *fission*, but Recanati explicitly proclaims himself against it. He is particularly worried that it would go against the empirical findings aforementioned, according to which visually keeping track of an object is shown to be cognitively more basic than the overly intellectual activity of parsing our perceptual contents into snapshots of distinct moments, i.e. that which I am attending to now, that which I was attending to some seconds ago etc. Besides these findings, it has been argued by some, with much approval, that dynamic thought is, in some sense, more fundamental than static thought, specially when demonstrative thought is concerned:

> We have to regard the static notion of 'having hold of an object at t^1' as essentially an abstraction from the dynamic notion of 'keeping track of an object from t^1 to t^2.' And the grasp, at t^1, of a thought of the kind suggested […] requires a subject to possess at t^1 an ability to keep track of a particular object over time. The capacities upon which certain kinds of thoughts rest can only be described in dynamic terms […] (Evans, 1981, p. 311)

Regardless of all that, I do not think there is any contradiction between the overall idea summarized by Evans' quotation and the thesis according to which demonstrative files are composite in the way I have suggested. I do agree that only conceptually sophisticated thinkers can pose questions about their perceptual episodes, such as whether one is hallucinating or being a victim of a surreptitious substitution. Thus, I think some thinkers, e.g. infants, are able to have demonstrative thoughts without being able to pose those questions. However, I am not making any claim about the cognitive fundamentality of static over dynamic thought (for all I know, any can be more fundamental). What I

am claiming, on the other hand, is that if a subject is cognitively sophisticated enough as to be able to perform *fission* on its demonstrative files, then those files must be based on a composite ER-relation, otherwise there would be no possibility of *fission* to begin with.

Analyzing the demonstrative ER-relation as composite in that way enables us to understand how we can perform *fission* in our demonstrative files. It also provides a neat explanation for why redeployment of a file outputs a new file that is merely W-DJCR to the original one. That is so because the original file was based on the static relation

$$\lambda x \lambda y (x \text{ is visually attending to } y \text{ at } t^1)$$

while the redeployed file is based on the distinct static relation

$$\lambda x \lambda y (x \text{ is visually attending to } y \text{ at } t^2 \text{ and } x \text{ was visually attending to } y \text{ from } t^1 \text{ to } t^2).$$

The latter part of the composite ER-relation contains the relation of the original simple ER-relation, thus, it ensures that both files will refer to the same object, if *both* refer at all. It is less clear how one could explain why redeployment leads to W-DJCR if one took the ER-relation of both files – at t^1 and t^2 - to be the dynamic relation of *x visually keeping track of y*, since then there would be no difference in the files' ER-relations to account for their referential divergence (one refers, the other does not).

Recanati himself (manuscript, p. 31) seems to gesture at this interpretation, where F^1 and F^2 respectively refer to a demonstrative file and its redeployment:

> The initial file F^1 stands in the W-DJCR relation to its successor F^2, since F^2 embodies a fallible presupposition of identity. File F^2 is, with respect to F^1, an 'inclusive', identity-dependent file, susceptible to type 3 cases of mistaken identity [i.e. one refers while the other does not]. So files F^1 and F^2 cannot stand in the strong DJCR relation; they can only instantiate W-DJCR.

The most natural reading of that quotation suggests that the redeployment of a demonstrative file will contain a presupposition of identity to its original deployment. In other words, the redeployment of a demonstrative file is more demanding than its original deployment, since it not only requires that there be an object demonstratively related to the thinker at the time of the tokening, but also that this object be the same as earlier. Putting these requirements inside the ER-relation of F^2 ensures that it will fail to refer if the object has disappeared or been substituted.

This story also seems to apply to the case of temporal indexicals. Imagine that a thinker utters "Today is a fine day" at a Monday 11:50pm (t^1). At Tuesday 00:00am (t^2), that thinker undergoes a drastic change of mood that results in her changing her mind about the belief she had expressed 10 minutes ago. However, she does not know that the day has already passed. With the intention of retracting from that utterance, she utters "Today is not a fine day, it is a terrible day!". Now, it is plausible that if this thinker changed her mind about her first utterance (and there appears to be no reason to say that she did not), the belief she expressed by her second utterance must contradict, in some relevant sense, the belief expressed by her first one. In other words, the cognitive value of the first belief must not be wholly distinct from the cognitive value of the retracted belief. The usual complication with that type of scenario is that the second utterance seems to refer to a distinct day than the first one, so, *per* (Factivity), they are cognitively distinct. Recanati's theory solves that problem in a clever way: the second utterance does not refer to Tuesday; it is empty due to violating a presupposition of coreference. That presupposition, his story would go on, is a constitutive part of the cognitive value of the second utterance: either it refers to the same day as the first utterance or it does not refer at all. That also entails that the two "Today" tokens in the story are W-DJCR, as opposed to DJCR. Thus, they are not cognitively identical *per se*, since that is the mark of DJCR tokens. On the other hand, they are not cognitively distinct *per se*, since that is the mark of independent tokens. They are cognitively related, and this relation is enough to ensure that the second utterance contradicts, in the proper sense, the first one[35].

Notice how Recanati's story also enables him to hold onto (Transparency). Our understanding of what makes a theory Non-Transparent so far has been that it allows thinkers having false beliefs about their own tokens' cognitive identity. Switching "cognitive identity" for "cognitive relation", we reach the more general principle that a theory is Transparent in regards to cognitive value if and only if it claims that thinkers can know, without recourse to experience, when their tokens are cognitively related, i.e. either W-DJCR or DJCR. If that principle is true, we can explain why the confused thinker believes that her two "Today" thoughts contradict each other. That is so because she knows, by virtue of introspection, that they are cognitively related[36].

---

[35] We could say that we track time analogously to how we visually track objects. Thus, we could characterize that thinker's TODAY file, in its original deployment at t^1, as being governed by the following ER-relation:

$\lambda x \lambda y (x$ is temporally attending to y at $t^1)$

while the redeployed TODAY file would be based on the distinct relation

$\lambda x \lambda y (x$ is temporally attending to y at t^2 and x was temporally attending to y at all instances before $t^2)$

[36] An interesting question is whether thinkers are able to distinguish, without recourse to experience, whether their tokens

A final consequence worth pointing out is that, according to my interpretation of Recanati, the reason why redeployment leads to W-DJCR is explained by means of *incremental conversion*. The second file is W-DJCR to the first one because it contains the first one's ER-relation plus something else. The adding of distinct ER-relations to a file is, *per* definition, the process of *incremental conversion*. Thus, *redeployment* is not a basic dynamic operation as it can be reduced to a more basic one.

4.4.2 CONVERSION

We can only think of a moment as NOW when that moment is presently related to us. Soon after we token a NOW thought, we are no longer able to think of that same moment through the same mode of presentation, we have to update it, or, in Recanati's preferred terminology: convert it.

Conversion, as we have previously seen, seems to be intrinsically related to the workings of memory. Thus, after a moment ceases to be thinkable as NOW, we convert its file into a memory file, which in English we usually express by means of "then". Whereas a NOW file's ER-relation is based on temporal tracking – i.e. $\lambda x \lambda y$(x is temporally attending to y) -, a THEN file's ER-relation is memory-based - i.e. $\lambda x \lambda y$(*x* remembers *y*). However, if these two files are based on wholly distinct ER-relations, how can there be any rational continuity between them?

To make that problem vivid, imagine that a thinker has, at t^1, a thought expressible by "Now is the coldest moment of the year". Some days later, at t^2, an even colder day happens to occur, and as a response to that, this thinker stops believing the thought she entertained at t^1. However, to achieve that change of mind, she cannot simply utter "Now is not the coldest moment of the year", since the thought expressed then would be cognitively distinct from her original thought – and change of mind about a certain token thought T^1 requires one to hold a contrary attitude to a token thought T^2 with identical (or, at least, related) cognitive value.

Recanati's solution to that problem is that one converts a NOW thought into a THEN thought as soon as the ER-relation grounding the first ceases to exist. Thus, when that thinker changes her mind at t^2, she will remember her THEN thought expressible by "Then was the coldest moment of the year" but will assent the contradictory THEN thought expressible by "Then was not the coldest moment of the year". Since these are clearly contradictory, there is a clear sense in which that thinker managed to change her mind.

However, I think there remains one problem that Recanati does not manage to solve. Why

---

are DJCR or merely W-DJCR. That is, whether the difference between DJCR and W-DJCR is transparent itself. I will leave that question unanswered.

should we concede that her post-conversion THEN thought is continuous with her original pre-conversion NOW thought? Since her NOW thought and her THEN thought share no ER-relation, there is no clear sense in which they are W-DJCR or DJCR. But, if they are neither W-DJCR nor DJCR, they are not cognitively related, thus, they are not 'the same thought' in the relevant sense with which we are concerned. Thus, while I grant that our thinker manages to change her mind about her THEN thought, I am yet to find an explanation for why that change of mind is also related to her original NOW thought.

I want to advance the following thesis: *conversion* ensures strong DJCR between its input indexical files and its output memory files because the latter are *mentally deferential* to the former. My biggest challenge will be explaining what exactly is mental deference - a process which I assimilate to the more familiar phenomenon of linguistic deference - and how it ensures DJCR between tokens associated to distinct modes of presentation. Thus, *contra* Recanati, I will claim that *mental deference* is a dynamic operation that guarantees DJCR between diachronic representations.

It is a familiar fact of natural language representation that speakers sometimes defer the meaning of their words to other people. Notice, for example, that the majority of people cannot differentiate instances of elm trees from beech trees even though they are competent users of those words. The natural way to explain that fact is by saying that, when these speakers utter tokens of 'elm' or 'beech', they are deferring the meaning of those words to the expert botanists of their community. The possibility of deference in natural language was historically associated with arguments in favor of semantic externalism (Burge 1979, Putnam 1975), since they seem to show that one can successfully refer to objects in the external world even when we have no substantial knowledge about those things or about the meaning of the words we use in referring to them. However, it is surprisingly easy to accommodate linguistic deference in a theory which acknowledges hyper-intensional semantic aspects. When speakers know very little about the referent of some of their words, the content they attach to these words contains explicit mention to the content, whichever it is, that other speakers associate with tokens of that same word. Imagine, for example, that someone called Jimbo eavesdrops on a couple of people talking about someone of that very same name. Jimbo could then go on to wonder "am I Jimbo?". If asked what he means by "Jimbo", a natural answer he could give is "whomever that couple was talking about". Similarly, an ignorant of botanic could think about elms without having any substantial knowledge about them because he takes 'elm' to refer to whichever kind of tree the wise people in his community refer to in their elm-talk. Without delving in this point any longer, I take it to be *prima facie* plausible that in linguistic deference we take some of our words' meanings as making explicit reference to what other people mean themselves by tokens of that same word.

There are many cases of linguistic deference. One can defer the meaning of a word to one

speaker, two speakers or the whole linguistic community around you. One can defer the meaning of a word to someone who does not know what he is speaking about, so linguistic deference can sometimes outputs empty words. One can also defer the meaning of a word to parties which disagree themselves about what that word means, so that deference can also output empty words when the deferred parties have meaningful words. Linguistic deference can also be a cooperative endeavor, in which parties defer to each other and no party has more entitlement to the meaning of a word than the other (Prosser, forthcoming). I want to focus in the simplest case of deference of all: when one creates a deferential use of a word from hearing someone else utter that word.

Thus, imagine that someone who is completely ignorant of Ancient History hears a professor say "Aristotle was Greek". Since that person is hearing that name for the first time in her life, she has absolutely no knowledge about its referent (except that he was Greek, something she also believes of other people). However, she then goes on to wonder about that person, who he was, whether he thought Greece was too hot in the summer etc. I claim that her grasp of her own "Aristotle" tokens is deferential, since we can easily imagine her saying that "Aristotle is whichever Greek guy the professor was talking about when he uttered 'Aristotle'". Thus, the reference-fixing conditions of her tokens would proceed via deference to the professor's tokens of the same name. In Mental File-talk, the ER-relation allowing her to acquire information about Aristotle is the relation of being deferentially related to the professor and his tokens of "Aristotle". Evidently, her grasp of her "Aristotle" tokens is completely distinct from the grasp the professor himself has of his "Aristotle" token, whatever that grasp consists in[37]. Regardless of that, it seems as if the ignorant thinker's tokens and the professor's token must necessarily corefer, if they refer at all.

Which kind of scenario would have to be the case for the ignorant thinker's "Aristotle" tokens to refer to something different from the professor's or even for one of the tokens refer while the other does not? It seems as if this is an impossible scenario! Whichever object the professor's tokens refer to, the deferred token will piggyback on that. If, on the other hand, the professor's token is empty (perhaps the professor is a lunatic), the deferred tokens will also be empty. If the ignorant thinker misheard and there existed no "Aristotle" token, the deferred token will again come out as empty. Still, there will be no instance of a Case 3 (typical of W-DJCR tokens) or Case 4 (typical of independent tokens). Our inability in coming up with a scenario in which a deferential token fails to corefer (if it refers at all) with its deferred token entitles us to conclude that linguistic deference ensures DJCR between input and output tokens, even though the ER-relations associated with each token by their

---

[37] I am not claiming the professor's grasp is a bundle of descriptive conditions nor anything of that sort. I am merely claiming it is not "whichever Greek guy the professor was talking about when he uttered 'Aristotle'".

owners are completely different.

One can say that deferential expressions, such as the ignorant thinker's "Aristotle", have their reference fixed by a token-reflexive rule that displays one interesting peculiarity. Instead of mentioning the deferential token itself, it mentions a completely independent token, the token to which deference is being made. A token T^1 of "Aristotle" made by the ignorant thinker, for example, refers to x if and only if a certain token T^2 (to which deference is being made) refers to x.

I contend that mental deference is alike linguistic deference in that precise respect, i.e. the rule which assigns reference to its tokens makes explicit reference to another token. Unlike indexical expressions, whose reference-fixing rules make reference to the same token to which it is assigning a reference, deferential expressions and memory demonstratives have their reference fixed via the semantic properties of other tokens. They piggyback, so to say, on the semantics of other representations.

There are, of course, important differences between deferential expressions and memory demonstratives. The former's function is to allow people who have very little knowledge about a token's referent to refer to it. Memory demonstratives are not about semantic ignorance, they are used to refer to things we have been acquainted in the past but are not anymore. Memory is obviously connected to perception in many ways. For example, it seems plausible that every memory demonstrative is grounded on some past perceptual demonstrative. That is, every singular memory we possess must have originated from some perceptual encounter we had in the past, even if, during that perceptual encounter, we did not explicitly token a perceptual demonstrative; what matters is that we could have tokened it if we had wanted it. The idea of a memory not grounded on a past perception hardly makes sense unless you are in a sci-fi convention. In Recanati's terminology, we can say that every memory demonstrative originates from the *conversion* of some past (perhaps merely possible) perceptual demonstrative.

However, there is a certain dependence between memory and perception which goes the other way round. If we do not *convert* a perceptual demonstrative into a memory demonstrative after the relevant perceptual episode is over, we lose the ability to entertain and retain the thought that it constituted.

What about the coreference relations between perceptual and memory tokens? I claim that the same reasons which have led us to claim that deferential expressions are DJCR with their deferred ones also leads us to claim that memory demonstrative tokens are DJCR with the perceptual tokens which they are a memory of. That is, since a memory demonstrative's reference is created by the *conversion* process, its reference is fixed via the perceptual token to which it is connected to. In other

words, a memory demonstrative M, created from converting a perceptual demonstrative P, refers to *x* if and only if P refers to *x*. If that is a correct representation of a memory demonstrative's reference-fixing rule, there just is no possibility in which these two tokens will fail to corefer (if they refer at all). If P fails to refer, so does M. If M is a faulty memory of a perceptual encounter which never existed, it does not refer and P is not real.

Let us now look into a couple of problems with this account (surely there are others). Firstly, there can be appropriate uses of a memory demonstrative that are not based on any past perceptual token. I have two distinct responses to that objection. My first response is claiming that a memory demonstrative can be created from *converting* a merely possible past perceptual token, that is, a token that could very well have existed but did not. That reply would continue by saying that we typically have memories of events in which we did not token any thoughts, but whose recollections are fine-grained enough for us to *suppose* that we did. Perhaps one does this by supposing that one is re-living the remembered event and supposing that one is tokening a demonstrative thought there. Another reply, which sounds simpler and more plausible, is that having a memory of an event entails that one perceptually represented it at the time it happened. Thus, even if there were no explicit and occurrent thoughts during the event, the reply would go on, there were tacit and implicit ones – and those can be brought to consciousness by means of reflection to give rise to memory demonstratives.

Secondly, some could claim that we can have singular memories without having any memory of a past perceptual token. Thus, even if someone's memory demonstrative M is based on one of her past perceptual demonstratives P, she may have forgot all about P's occurrences and still possess M, thus, her grasp of M cannot be explained by her having in mind some rule mentioning P's tokens. I am not sure whether this thesis is true. Anyone in possession of a singular memory will know that their memory demonstratives refer to whomever they were in perceptual contact with during the time of the remembered event. Even someone completely ignorant of all philosophical matters will be able to realize that "that man" in "that man was wearing his pajamas at work" refers to the man she saw some day at work. Since the act of seeing entails a perceptual contact, that thinker has more than enough resources from which to create a memory demonstrative in accordance with the story I am suggesting.

Without going any further on those matters, let me just claim that, if anything like this story is true, then, *contra* Recanati, there can be diachronic tokens that are DJCR with each other. Memory demonstratives (and the perceptual demonstratives which they are a memory of) are an example of this. Thus, we reach the interesting conclusion that some tokens can bear the strongest rational relation to each other (DJCR) without them being associated with the same reference-fixing conditions. This shows that the notion of cognitive value, since it is associated with DJCR, might not perfectly track the

identity of modes of presentation. At this point, we reach the predicament of having to decide whether the identity conditions of propositional attitudes, such as beliefs, should supervene on the identity of the modes of presentation therein contained or their cognitive value. It is possible that two cognitively identical beliefs do not present their referents in the same way, e.g. my perceptual belief [at t^1] that *that man is wearing black* and my subsequent memory belief [at t^2] that *that man was wearing black*. This is a particularly worrying predicament for the Fregean philosopher, since she will now have to decide whether the identity conditions of Senses should track the instantiation of DJCR or the modes of presentation (i.e. reference-fixing rules) associated. Anyway she decides to go, what matters is that a semantic theory needs to carve semantic space in three levels, ordered by increasing granularity: intensional level, cognitive level and mode of presentation (henceforth MOP) level .

The intensional level seems to supervene on both the cognitive and MOP levels, i.e. there can be no cognitive or MOP identity without an intensional identity. The cognitive level, for its turn, seems interestingly independent of the MOP level. Since there can be MOP-distinct tokens about which it would be irrational to wonder whether they corefer (if they refer at all), as in the case of deference or memory, there can be cognitive identity without MOP-identity. However, I do not think the converse is true, that is, I do not think that there can be MOP-identical tokens which are cognitively distinct. It seems that, if one understands two MOP-identical tokens, one must be able to see that this so and infer that they should corefer (if they refer at all).

To pursue these issues with more precision, one would need to advance a theory of modes of presentation and their identity conditions. One question that this philosopher would need to answer is, are logically equivalent MOPs the same? That is, is a name for the antecessor of 2 MOP-identical with a name for one half of 2? Another question is, are semantically equivalent MOPs the same? That is, is a name for the tallest lawyer MOP-identical with a name for the tallest attorney? Unfortunately, I am not sure how to answer these questions.

4.5 BACK TO ALICE, BOB AND CAROL: MEMORY AND BELIEF RETENTION

According to Recanati's theory, dynamic operations only guarantee W-DJCR between their inputs and outputs. *Contra* him, I have argued that *conversion*, i.e. the transformation of perceptual demonstratives into memory demonstratives, guarantees DJCR between its input and output tokens.

In one passage (forthcoming, p. 17), Recanati explicitly assumes that the dynamic operation which lead us from "Today" beliefs to "Yesterday" beliefs is that of *conversion*. I will now claim that it cannot be so for two reasons. Firstly, *conversion* infallibly guarantees that its input token will corefer (if it refers at all) with its output token, and that is not the case with the *today/yesterday* transformation.

Secondly, *conversion* is a process that converts demonstrative thoughts into memory thoughts. However, even if "Today" thoughts can be characterized as demonstrative in a certain sense, I do not agree that "Yesterday" thoughts are purely memory-based.

About the first reason, we have seen in the previous subsection that memory demonstratives are guaranteed to corefer (if they refer at all) with demonstrative tokens which they are a memory of. Furthermore, that guarantee is *a priori*, since there mere understanding of the reference-fixing conditions of both tokens produce knowledge of their conditional coreference. The situation here is analogous to paradigmatic cases of DJCR, such as the relation between the singular tokens in "Cicero is tautologically identical to Cicero", "Marcone is a butcher but he is a vegetarian" etc. These cases have one trait in common: the conditional coreference of their singular tokens is infallible.

Now, notice that the *today/yesterday* transformation necessarily lacks this infallibility property, that is, it is always empirically possible that one has lost track of time to the effect that one's "yesterday" tokens will fail to corefer with one's past "today" tokens. Even in cases where we are rationally entitled to assume that some "yesterday" token corefers with some "today" token, we cannot rule out, by purely *a priori* means, the possibility that we are living some sort of Bob's scenario. In summary, knowledge of the coreference between a "today" and a "yesterday" token depend on knowledge about the context in which they were uttered; since that is only knowable *a posteriori*, knowledge of the coreference between some "today" and "yesterday" token is also *a posteriori*. That is enough to show that, whichever dynamic operation we transforms "Today" beliefs into "Yesterday" beliefs, it is not the same one we use for converting demonstrative beliefs into memory beliefs and I intend to reserve the label of *conversion* for the latter one.

The second reason why I do not think *conversion* is at play in the *today/yesterday* transformation is that "Yesterday" beliefs do not seem to be purely memory beliefs in the required sense. As of proof of that, notice that one may *convert* a "Today" belief into a memory belief without ever holding a "Yesterday" belief. Take Carol, for example, the thinker who had many reasons to suspect that she had slept for more than one day since Monday. If she were fully rational in re-expressing her previously had "Today" belief, she would not choose to do it by means of "Yesterday", but by means of a memory demonstrative that, in English, we usually express by means of "That", i.e. "That [remembered] day was a wonderful day". As a rule, anyone who loses track of time will be able to re-express one's past temporally indexical beliefs by means of a memory demonstratives. That stablishes that the retention of temporally indexical beliefs only requires their *conversion* into memory beliefs, even if one has completely lost track of time. Additionally, I take that to stablish that a memory belief, such as that expressible by "That [remembered] day was a wonderful day", is cognitively

distinct from a "Yesterday" belief, such as that expressible by "Yesterday was a wonderful day". That is so since anyone could simultaneously believe one while rationally disbelieving the other.

In summary, I claim that, to retain one's indexical beliefs one must *convert* them into memory beliefs. That *conversion* does not presuppose having kept track of time, since even the most disoriented of all thinkers can successfully retain his memories. Thus, I claim that the focus that has been given on the issue of temporal tracking is misguided, since belief retention must be explained by memory, and not by having true beliefs about the contexts on which certain thoughts were produced. As Branquinho (2000) notices, the paradigmatic thinker who is not able to retain her indexical beliefs is not one, such as Bob, who loses track of time, but an amnesiac one who cannot retain her own memories.

If one agrees with what I have just said, one has to make a certain theoretical decision. Either there is a dynamic operation, distinct from *conversion*, corresponding to the *today/yesterday* transformation or there is no dynamic operation in that vicinity at all.

The first option, which seems to be implicit in Branquinho (2000) and Perry (1996), entails that there is some intrinsic continuity between, e.g. Alice's "Today" and "Yesterday" belief, even though there is an even deeper continuity between the former and a memory belief expressible by "That [remembered] day was a wonderful day". If that were the case, there would be two distinct ways to retain a "Today" belief, either by *converting* it into a memory belief or by transforming it into a "Yesterday" one. Since, as I have claimed, those output beliefs are cognitively distinct, the two types of retention would lead one to cognitively distinct beliefs, thus, distinct beliefs in the relevant sense. I think it is clear that there is something awry with that theory of belief retention. The idea that we can retain a belief in multiple manners that are not cognitively convergent turn belief retention into a much too idiosyncratic process.

The second option, which I am much more sympathetic to, claims that there is no retention, in the sense of internal continuity, between "Today" and "Yesterday" beliefs. The reason is that their coreference relations are necessarily an *a posteriori* matter, whereas I claim that internal continuity presupposes some kind of *a priori* guarantee of referential convergence. Thus, I want to reject the claim that there is DJCR between "Today" and "Yesterday" tokens, even if the subject has successfully managed to keep track of time. However, to defend that position, I need to go against Frege's (1997, p. 296) influential remark to the effect that "if someone wants to say the same today as he expressed yesterday using the word 'today', he must replace that word with 'yesterday'". Even more pressingly, I need to explain why thinkers, such as Alice and Bob, believe that, by means of "Yesterday was a wonderful day", they are expressing the exact same belief they had previously expressed by means of

"Today is a wonderful day". That is, it seems that I am contradicting the transparency of our own thoughts by claiming that thinkers can believe they are re-expressing their beliefs while in fact expressing a cognitively distinct one.

Let me provide an explanation for that, though admittedly it will be only a sketch of an explanation. We normally reason under many background assumptions, one of which is the assumption that we successfully have tracked the passage of time. Since that assumption is rarely contested, we tacitly assume that some identities, such as that our "Today" tokens uttered at a certain day are coferential with our later "Yesterday" tokens, are true. However, that tacit assumption is qualitatively distinct from our presuppositions of coreference between DJCR tokens, such as our presupposition that the two "Cicero" tokens in our belief that Cicero is identical to Cicero are coreferential (if they refer at all). The latter assumption is not up to rational revision, while the former is. By means of analogy, our assumption that our present "Yesterday" tokens corefer with some past "Today" tokens is like our assumption that "Pelé" is coreferential with "Edson Arantes do Nascimento", i.e. even though they are obvious, they are grounded in empirical evidences and up to rational revision. In contrast to that, our beliefs in the conditional coreference of paradigmatic DJCR tokens are never grounded in empirical evidence; these are, not only obvious, but also *a priori*. Thus, I claim that, when someone reasons as if some past "Today" token were obviously coreferential with some present "Yesterday" token, one is, in reality, relying on a tacit *a posteriori* premise of identity. Since this premise is rarely contested, it often goes unnoticed. However, that premise is there and could be made explicit by reflection on the epistemological grounds of our inferences and beliefs.

For the sake of illustration, imagine that Bob realizes he has been subject to a drug while sleeping. It is plausible that he would then say that by his "Yesterday" utterance, what he really meant was to refer to that (remembered) day, and not to the previous day. That suggests two natural ways to rationalize Bob. The first way would be claiming that he has not used the expression "Yesterday" with its usual meaning, since he expressed a cognitive value that is proper to memory demonstratives. This would be equivalent to claiming that Bob has used "Yesterday" metaphorically or outside its ordinary meaning. The second way would be claiming that, in his previous reasoning, he was relying on the tacit premise that "Yesterday = the day I remember".

One important consequence of these suggestions is that proper belief retention, that is, retention of the strongest kind, necessarily depends on converting one's indexical representations into memory representations. However, memory representations are not indexical in the same way as paradigmatic temporal indexicals or demonstratives are. Their reference-fixing conditions are not rules picking out a feature of the context of utterance, as usually is the case of indexical tokens. Instead, a

memory demonstrative picks out its reference via another distinct token. Thus, they are more similar to deferential tokens, instead of indexical ones. This entails that proper retention of one's indexical attitudes presuppose converting them to attitudes that are not themselves indexical. Whether this should be seen as a surprising consequence or not is an interesting issue to be researched upon.

The picture we arrive by accepting these suggestions is one that requires updating our definition of DJCR. Since properly DJCR tokens can never end up failing to corefer (if they refer at all), we can scrap the mention to "rational entitlement" in its definition in favor of the relation of *a priori* knowledge:

(**DJCR²**) Two tokens A and B are DJCR if and only if one who understands them is able to know *a priori* that they corefer if they refer at all.

Naturally, we would need more discussion in order to assess whether *de jure* coreference should really be defined by means of *a priori* knowledge, instead of the weaker epistemic relation of rational entitlement. We would, for example, have to consider the case of perceptual tracking (as opposed to temporal tracking), where it is *prima facie* less plausible to claim that we rely on tacit premises of identity in our reasoning (Campbell, 1987). All in all, I take myself to have at least motivated the thesis that, as far as temporal tracking and temporal indexicals go, DJCR should be strengthened as to allow infallible knowledge of conditional coreference.

5. CONCLUSION

In chapters 2, I have defended a series of principles which jointly give rise to – what I take to be – a broadly Fregean theory of singular content. The first principles I have introduced were (A1) and (Clausalism):

(A1) two cointensional representations may differ semantically,

(Clausalism) All cognitive differences between cointensional representations is due to cognitive differences between its referential tokens,

According to the first, the cognitive differences between cointensional representations are to be explained semantically, e.g. *via* postulation of a finer-grained semantic value. According to the second, these cognitive differences are to be located in the singular tokens of these representations. In a first moment, I have tried to flesh out the notion of cognitive value in terms of the Token Doxastic Test:

(**Tokens Doxastic Test**) If two representations differ solely by the substitution of coreferential tokens and one may rationally bear conflicting attitudes towards them, then these coreferential tokens are cognitively distinct.

However, I have later argued that the explanatory power of that principle can be enhanced by acknowledging the relation of *de jure* coreference:

(DJCR) Two tokens A and B are DJCR if and only if one who understands them is rationally entitled to believe that they corefer if they refer at all.

As I have showed, it is easy to prove that any purported cognitive discrepancy between referential tokens can be reduced to the instantiation (or lack thereby) of DJCR by them:

(A2) Two tokens A and B are cognitively identical if and only if A and B are DJCR.

I have argued that the philosophical problem with which Frege (1949) was concerned in his seminal *Über Sinn und Bedeutung* - that which has been called "Frege's Puzzle" - was, indeed, founded

upon the recognition of DJCR. Thus, Frege's Puzzle should be reformulated as a puzzlement about the source of that relation:

(Frege's Puzzle) In virtue of what are some tokens DJCR (such that their identity is trivial) while others are independent (such that their identity is informative)?

Since the acknowledgement of DJCR does not presuppose any substantial theory of singular content, I have used it to show how the exposition of Frege's Puzzle is not vacuous nor circular, in contrast to what Glezakos (2009) has argued. Furthermore, DJCR can be used to justify Frege's own historical solution of the puzzle, i.e. the introduction of Senses:

(A3) Two tokens A and B have the same Sense if and only if A and B are DJCR.

This reframing of Frege's original argument in favor of Senses remind us of his use of principles of abstraction to argue for the existence of numbers in the *Grundlagen* (1953). Analogously to the role which equinumerosity plays in drawing the boundaries of the concept of a natural number, the instantiation (or lack thereby) of DJCR can be used to determine whether two tokens have (or not) the same Sense.

In chapter 3, I have tackled two objections purporting to claim that DJCR is a non-transitive relation. If these objections were true, something its proponent Pinillos (2011) would very much appreciate, they would show that DJCR cannot be used to ground the identity conditions of Senses, since identity is an equivalence relation. Contrary to that, I have showed how Pinillo's purported examples of non-transitivity were really pointing to a distinct coreference relation, which I called, following Recanati (2015), "weak *de jure* coreference":

(W-DJCR) Two tokens A and B are W-DJCR if and only if one who understands them is rationally entitled to believe that either they corefer (if they refer at all) or that one of them refers while the other does not.

W-DJCR is weaker than proper DJCR since it allows that one of its *relata* my fail to refer, while the other does not. By defending the thesis that Senses can enter into multiple logical operations, such as conjunction, I have showed that they can properly explain why some tokens are merely weakly *de jure* coreferential:

(A5) Any two Senses α and β can be composed to form the composite Sense α and β

(A6) Two tokens A and B are W-DJCR if and only if either the Sense of A has a part in common with the Sense of B or the Sense of B has a part in common to the Sense of A.

Furthermore, I have argued that the Fregean philosopher is more well-suited to explain the cognitive differences between W-DJCR and DJCR tokens than the relationist one. My argument was that Pinillos' (2011) own relationist theory, in contrast to the Fregean one I have defended, does not have enough theoretical resources to explain

(Cognitive difference between DJCR and W-DJCR tokens) For any two DJCR tokens, but never for merely W-DJCR tokens, we are entitled to make both substitution inferences with them.

The last part of chapter 3 is concerned with a distinct type of counterexamples suggested by Pinillos. This time, I have argued that his counterexamples are based on a mistaken premise, i.e. that one cannot truly ascribe a contradictory belief to a rational subject. Against that, I have argued that, if Pinillos' examples are read as *de re* ascriptions, they present no counterexample to the transitivity of DJCR. After that, I have briefly discussed the issue of the truth-conditions of propositional attitudes, suggesting that a Fregean should abandon:

(Intensionality) There is never any fact about truth-conditions which is not reflected in intensional content/reference.

As long as one abandons (Intensionality), I have claimed, one can easily explain why cointensional attitude ascriptions may differ in truth-value without entailing that tokens shift their referents in special representational environments.

Finally, the appendix *Levels of Identity/Distinctness between Identity Representations* demonstrates how, after the acknowledgement of DJCR, four different degrees of semantic identity for identity representations can be defined, each pointing towards properties (of varying strength) that two identities may instantiate.

Chapter 4 departs from the previous discussion with the hope of applying what we have

learned about DJCR to the case of indexical representations; more particularly, those that are diachronic, i.e. temporally distant. I consider the proposal of Evans (1981) and dismiss it because it denies us a transparent access to the identity conditions of our thoughts, i.e. the instantiation (or lack thereby) of DJCR between its tokens. More interestingly, I argue that Bozickovic's (2015) theory, originally intended to be an improvement over Evans', should also be dismissed for similar reasons. Even though Bozickovic enforces the principle of transitivity, he lets go of the principle of factivity. However, as I argue, letting go of one of these principles in favor of the other incurs in no significant theoretical gain for a theory.

I then consider Recanati's (2012, 2015, forthcoming) thesis that there can only be W-DJCR between diachronical indexical tokens. While I agree with most of his claims, I suggest two places in which his proposal should be amended. Firstly, I claim that the redeployment of an indexical token should be seen as involving incremental conversion of the original deployment. Otherwise, there will be no straightforward way to explain how one could rationally come to wonder about the coreference of a token and its redeployment. Secondly, I claim that there can be DJCR between diachronic indexical tokens in the case of conversion, where a perceptual demonstrative is transformed into a memorial one. Interestingly, that realization leads us to conclude that there can be cognitively identical tokens associated to distinct modes of presentation. Since memory demonstratives are more alike deferential expressions than indexical ones, this realization also suggests that the most suitable way to retain our indexical beliefs, i.e. converting them into memory beliefs, involves detaching them from their originally indexicality.

Finally, I end chapter 4 by defending two points. One, very general, the other, very specific. The specific point is that "Yesterday" tokens are never DJCR with past "Today" tokens, even if the subject has correctly kept track of time. I defend that claim by showing that proper belief retention of a "Today" belief should be done by converting it into a memory belief, i.e. one that we would express, in English, by "that (remembered) day". Furthermore, I argue that our commonsensical intuitions that we are samesaying when we re-express a "Today" belief by means of a "Yesterday" token hides the fact that we are usually using "Yesterday" in a non-literal sense, i.e. we are using "Yesterday" as if it meant "that (remembered) day".

My general, and, admittedly, more tentative, point is that, if my arguments are correct, there are no cases in which two DJCR tokens fail to corefer (if they refer at all). That means that it may be possible to strengthen the definition of DJCR even more by scrapping the concept of "rational entitlement" in favor of "*a priori* knowledge":

(DJCR²) Two tokens A and B are DJCR if and only if one who understands them is able to know a priori that they corefer if they refer at all.

REFERENCES

ALMOG, Joseph. 2008. Frege puzzles? Journal of Philosophical Logic 37 (6):549 - 574.

BOCCARDI, Emiliano. 2012. Frege's Puzzle: Much Ado About Nothing? In Revista Filosofía Univ. Costa Rica, 53 (136 Extraordinary), 133-146.

BOGHOSSIAN, P. A. The transparency of mental content. Philosophical perspectives, pages 33–50, 1994.

BOZICKOVIC, Vojislav. 2015. Belief Retention: A Fregean Account. Erkenntnis 80 (3):477-486.

BRANQUINHO, João. 2000. On the individuation of Fregean Propositions. The Proceedings of the

BURGE, Tyler. 2005. *Truth, Thought, Reason: Essays on Frege*. Oxford University Press.

CAMPBELL, John. 1987. Is sense transparent? In Proceedings of the Aristotelian Society, volume 88, pages 273–292.

CAPPELEN, Herman. 2012. *Philosophy Without Intuitions*. Oxford University Press.

CARNAP, Rudolf. 1947. *Meaning and Necessity*. University of Chicago Press.

CASTANEDA, H. N.; HART, J. G. & KAPITAN, T. (eds.) 1999. *The Phenomeno-Logic of the I: Essays on Self-Consciousness*. Indiana University Press.

DAVIDSON, Donald. 1968. On saying that. Synthese 19 (1-2):130-146.

DICKIE, Imogen & RATTAN, Gurpreet. 2010. Sense, Communication, and Rational Engagement. Dialectica 64(2):131-151.

DORR, Cian. 2014. Transparency and the Context-Sensitivity of Attitude Reports. In Manuel García-Carpintero & Genoveva Martí (eds.), *Empty Representations: Reference and Non-existence*. Oxford University Press 25-66.

EVANS, Gareth. 1981. Understanding demonstratives. In Herman Parret (ed.), *Meaning and Understanding*. Clarendon Press 280--304.

FARIA, Paulo. 2009. Unsafe reasoning: a survey. Doispontos 6 (2):185-20.

FIENGO, Robert & MAY, Robert. 2006. *De Lingua Belief*. Cambridge MA: Bradford Book/MIT Press.

FINE, Kit. 2007. *Semantic Relationism*. Blackwell Pub.

FODOR, Jerry A. 1975. *The Language of Thought*. Harvard University Press.

FODOR, Jerry A. 2008. *Lot 2: The Language of Thought Revisited*. Oxford University Press.

FREGE, Gottlob & BEANEY, Michael (eds.) 1997. *The Frege Reader*. Blackwell Publishers.

FREGE, Gottlob. 1948. Sense and reference. Philosophical Review, 57 (3):209-230.

FREGE, Gottlob. 1949. On sense and nominatum. In Herbert Feigl & Wilfrid Sellars (eds.), Readings in Philosophical Analysis. Appleton-Century-Crofts p. 85-102.

FREGE, Gottlob. 1953. *The Foundations of Arithmetic*. Evanston, Ill., Northwestern University Press.

GERKEN, Mikkel. 2011. Conceptual Equivocation and Warrant by Reasoning. Australasian Journal of Philosophy 89 (3):381-400.

GERKEN, Mikkel. 2013. *Epistemic Reasoning and the Mental*. Palgrave Macmillan (Innovations in Philosophy).

GLEZAKOS, Stavroula. 2009. Can Frege pose Frege's puzzle? In Joseph Almog & Paolo Leonardi (eds.), *The Philosophy of David Kaplan*. Oxford University Press, p. 202.

HAHN, Martin. 1995. The Frege puzzle one more time. In Petr Kotatko & John Biro (eds.), *Frege: Sense and Reference One Hundred Years Later*. Kluwer 169-183.

HECK Jr, Richard G. 2012. Solving Frege's Puzzle. Journal of Philosophy 109 (1/2):132-174.

KAPLAN, David. 1968. Quantifying in. Synthese 19 (1-2):178-214.

KAPLAN, David. 1990. Words. Aristotelian Society Supplementary Volume 64:93-119.

KRIPKE, Saul A. 1979. A puzzle about belief. In A. Margalit (ed.), *Meaning and Use*. Reidel 239--83.

LOAR, Brian. 1988. Social content and psychological content. In Robert H. Grimm & D. D. Merrill (eds.), *Contents of Thought*. University of Arizona Press

NINAN, Dilip (2014). On Recanati's Mental Files. Inquiry 58 (4):368-377.

ONOFRI, Andrea. 2015. Mental Files and Rational Inferences. Inquiry 58 (4):378-392.

PERRY, John. 1993. *The Problem of the Essential Indexical: And Other Essays*. Oxford University Press.

PERRY, John. 1996. Rip Van winkle and other characters. European Review of Philosophy 2:13-39.

PINILLOS, N. Angel. 2011. Coreference and meaning. Philosophical Studies, 154(2):301–324.

PRIOR, A. N. 1959. Thank Goodness That's over. Philosophy 34 (128):12 - 17.

PRYOR, James. 2016. *De Jure Codesignation*. Available at: http://www.jimpryor.net/research/papers/DeJure.pdf Accessed at: March 20th, 2016.

PYLYSHYN, Zenon W. 2007. *Things and Places: How the Mind Connects with the World*. The MIT Press.

RECANATI, François. 2012. *Mental files*. Oxford University Press.

RECANATI, François. 2015. Replies to the Papers in the issue "Recanati on Mental Files". Inquiry, 58 (4):408-437.

RECANATI, François. 2016. Indexical Thought: the Communication Problem. In S. Torre and M. Garcia-Carpintero (eds.) *About Oneself*. Oxford University Press.

RECANATI, François. Forthcoming. Cognitive dynamics. To appear in K. Korta and M. Ponte (eds.) *Reference and Representation in Language and Thought*.

SALMON, Nathan U. 1986. *Frege's Puzzle*. Ridgeview.

SCHELLENBERG, Susanna. 2012. Sameness of Fregean sense. Synthese 189 (1):163-175.

SCHIFFER, Stephen. 1978. The basis of reference. Erkenntnis 13 (1):171--206.

SCHOLL, Brian. 2001. Objects and attention: the state of the art. Cognition 80 (1-2):1-46.

SOAMES, Scott. 2002. *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford University Press.

STALNAKER, Robert. 2008. Our Knowledge of the Internal World. Oxford University Press.

TAYLOR, Kenneth. 2015. Names as Devices of Explicit Co-reference. Erkenntnis 80(S2):235-262. Twentieth World Congress of Philosophy, 2000:17–27.

WETTSTEIN, Howard. 1989. Turning the tables on Frege or how is it that "Hesperus is Hesperus" is trivial? Philosophical Perspectives 3:317-339.