Proper Names
A Cognitive-Philosophical Study

by

Eduardo García-Ramírez

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
in The University of Michigan
2010

Doctoral Committee:
        Assistant Professor Eric P. Swanson, Chair
        Professor Marilyn J. Shatz
        Professor Richmond H. Thomason
        Associate Professor Andrew M. Egan

Wie seltsam, wenn sich die logic mit einer 'idealen' Sprache
befaßte und nicht mit *unserer*.

— Ludwig Wittgenstein
Philosophische Bemerkungen

. . .

A conspicuous difference between old epistemology and the
epistemological enterprise in this new psychological setting is
that we can now make free use of empirical psychology.

— W.V.O. Quine
Epistemology Naturalized

To my family

— in loving memory

Acknowledgments

This dissertation would not have been possible without the help and support of my committee: Eric Swanson, Marilyn Shatz, Andy Egan, and Rich Thomason. My work has been greatly influenced by their guidance. Each one of them has, in her / his own way, helped me become a much better philosopher.

Chapters 2, 3, 4, and 5 have benefited from presentations at different venues. Thanks are owed to audiences at the Proper Names Reading Group (PNRG) at UofM, the graduate seminar on Psycholinguistics at UofM, the Philosophy of Language Seminar at IIF's-UNAM, the second annual AAMPh conference at CUNY, the faculty seminar of the Logic and Philosophy of Science Department at UCM-Madrid, the Vortragsreihe at the Ruhr-Universität Bochum, the workshop on Cognitive Perspectives on Mind and Language at IIF's-UNAM, and the first annual EFA conference at UBA-Buenos Aires. Special thanks are owed to Lina Jansson, Sam Liao, Ivan Mayerhoffer, Nate Charlow, Maite Ezcurdia, Mario Gómez-Torrente, Axel Barceló, Ángeles Eraña, Agustín Rayo, Rasmus Winther, Carla Merino, Carlos Montemayor, Raúl Saucedo, Ricardo Mena, Luis Fernández, Albert Newen, Ulrike Pompe, Christoph Michel, Isidora Stojanovich, and Edouard Machery. Special thanks are also owed to Lenny Clapp and Francois Recanati for their detailed comments on previous versions of chapter 4, and to Marilyn Shatz with whom I coauthored chapter 2. Thanks are owed as well to Dustin Tucker and Molly Mahony for their editorial help.

This dissertation also benefited from two term-long visits to MIT-Philosophy and to IIF's-UNAM. Thanks to Agustín Rayo and Guillermo Hurtado for making them possible.

My philosophical development has also benefited enormously from the wisdom of my peers. I feel fortunate for having had the chance to work with the class of 2005. Substantial thanks are owed to Alex Plakias, Lina Jansson, Wendy Tng, Sam Liao, Michael Docherty, Neil Mehta, Ian Flora, and David Plunkett. I am also fortunate for being part of a Department with such an excellent group of graduate students. Thanks are owed to all.

My research has improved substantially thanks to the amazing work of our librarian. Thanks are owed to Molly Mahony for keeping me up to date with respect to proper names. Thanks are also owed to Linda Shultes, and to all the philosophy staff, for turning the PhD

into a less daunting, more liveable, task.

Not everything has been nice during the past few years. I want to thank the faculty and staff of the Philosophy Department at UofM for their extraordinary support during the horrible times. Thanks are owed as well to Abraham Arceo, Juan Espíndola, Jonathan DeVore, Luciana Meli, Bruno Velazquez, Moisés Vaca, and Marco Rigau for their unyielding support.

Finally, I want to thank Catalina for keeping up with a partner-unfriendly academic lifestyle and, most importantly, for keeping me alive.

Table of Contents

List of Tables

List of Figures

Abstract

Proper Names appear at the heart of several debates in philosophy and the cognitive sciences. These include *reference*, *intentionality*, and the nature of *belief* as well as *language acquisition*, *cognitive development*, and *memory*. This dissertation follows a cognitive approach to the philosophical problems posed by proper names. It puts an emphasis on adequately describing the actual cognitive abilities that allow humans to understand and use proper names. To achieve this goal I use the evidence obtained by cognitive and neuropsychology as well as psycholinguistics. The result is an empirically informed and integrative theory of reference, language use and intentionality, and some metaphysics and epistemology thereof.

Chapter 2 argues against descriptivism and partly develops a theory according to which an understanding of *reference* is prelinguistic and requires no mediating descriptive intensions. The theory, which is consistent with a referentialist semantics for names, is supported by empirical data from lexical, cognitive and neuropsychological research on the acquisition, understanding and processing of proper names. If correct, this theory has important consequences for the debate concerning the possibility of singular thought.

Chapter 3 describes the relation between proper name use and the structure of belief by reflecting on the intimate connection between the puzzles of *informativeness* and *substitution-failure*. Chapter 4 offers a psychological and linguistic resolution of the puzzles. It offers an account of the psychological architecture involved in proper name processing and shows how it interacts with a linguistic account of how proper names are used in tandem with predicate meaning-transfer.

Chapter 5 offers a theory of empty names by using some cognitive-general, non-linguistic, resources. A special kind of mental representational state (i.e., EDU-attitudes) and its content (i.e., EDUs) is described and applied. By solving the problems posed by the ordinary use of empty names, the theory helps illuminate issues in metaphysics, ontology, intentionality, and fictional imaginings. Chapter 5 also offers a metalinguistic account of negative existential assertions (e.g., utterances of the form 'X doesn't exist'). This account helps us avoid a common dilemma between a controversial ontology and an unsatisfactory linguistic theory.

Chapter 1

Introduction: Why should we care about proper names?

Why should a philosopher care about proper names? I see proper names as located at two different thresholds: first, between the world of our thoughts, beliefs and desires and the world of our actions, institutions, and practice; and second, between the linguistic and the non-linguistic, between the world of meanings and truth-values and the world of representations and imaginings. Thus, proper names turn out to be important for several reasons.

On the one hand, a proper understanding of the ordinary use of proper names illuminates several cognitive phenomena such as language acquisition and use, as well as the nature of thought and belief (intentionality). On the other hand, there is no understanding of proper names without a proper understanding of reference. Thus, proper names also teach us about how our thoughts and assertions relate to the world, thereby teaching us about the latter by telling us what the former are about.

We have thoughts about how the world is, but we also like to think about how it should be. So, presumably, a proper understanding of reference and intentionality should *help* us understand not only our claims to knowledge but also our value judgments. I think a satisfactory account of proper names sheds *some* light over these issues as well (see Shatz 1994 and 2008).

To address the central issues I have decided to follow a more recent methodology: what has come to be known as "cognitively oriented" philosophical theorizing (see Leslie 2007 and 2008). This means that I will be mostly worried about how it is that human beings actually manage to acquire, understand, and use a given natural language and, in particular, proper names. Thus, my main goal is *not* to offer a formal semantic model for proper names, one that we may then associate with compositional rules in order to recursively assign truth-conditions to all possible sentences involving proper names.

Throughout, I am interested in finding a way to better understand the actual cognitive abilities required for natural language speakers to do what they do when using proper names. Since these abilities are responsible for our understanding of proper names, providing an account of them should allow us to understand the contribution made by a proper name to the content of a thought. I believe this approach is complementary to the more traditional one; yet the methodologies are clearly distinct (for a recent contrasting approach, see Cumming 2008). A lot of work has been done on the logic of proper names. I think it is time to start working on getting the psychology straight. With this dissertation I hope to contribute to that part of the philosophical project.

1.1   Reference

Proper names are perhaps the most paradigmatic devices of reference. They are independent of perceptual characteristics, universal across languages, and primitive for all of them (see Hall, 1999). Furthermore, they are learned much before (4.5 - 6 months of age) any other terms (some common nouns appear at 9-10 months and adjectives at 18), suggesting that proper names are among the building blocks of language development.

There are two broad views about reference for proper names (see Reimer 2003). On the one hand, there are those who think that reference is not mediated. On this view, the contribution of a proper name to the content of a thought just is its referent. On the other hand, there are those who think that reference is mediated. Those who like this latter view have generally proposed what is known as "sense" theories of reference, where a sense is supposed to play the mediating role between the referent and the thought. This view is associated with description theories of reference since it is common to identify senses with the content of definite descriptions.[1]

In chapter 2 we argue against the view according to which reference for proper names is mediated by definite descriptions.[2] We offer an empirical assessment of description theories of names by looking at evidence on lexical and cognitive development, memory, and aphasia. We show that description theories demand much more cognitive resources than what the data suggest and so they lack empirical support.

In this chapter we also offer a preliminary developmental account of reference for proper

---

[1]Not all description theories of names count as description theories of reference. Russell 1905 argues that names are not referential devices but disguised descriptions, which he analyzes in terms of quantifiers. On this view, reference is not mediated by a description, but names do have a descriptive content as their semantic value.

[2]Chapter 2 was coauthored with Marilyn Shatz. A version of it will appear in *Mind & Language*.

names. On this view, an understanding of reference is the result of a prelinguistic prepared-ness on the side of subjects to understand that a given object (e.g., the string of sounds 'Jon') can be the sign of something else (e.g., Jon). We offer some evidence on behalf of this account.

As a result, this chapter offers indirect support for the view that there is direct reference and, thus, for the view that the objects to which we refer can constitute the content of our thoughts. The issue of whether there are some such thoughts—also known as "singular thoughts"— is still hotly debated among philosophers (see Recanati forthcoming). I believe the arguments in chapter 2 can substantially illuminate this debate.

Mallon, Machery, Nichols, and Stich 2009 argue against drawing philosophically sig-nificant conclusions out of theories of reference. This is so because, the authors claim, "intuitions play a central role in establishing theories of reference, and recent cross-cultural work suggests that intuitions about reference vary across cultures and between individuals in a culture" [p.332]. The account of reference we present in chapter 2 is based on evi-dence from lexical and cognitive development as well as on research from neurophysiology. Intuitions do not play a central role in it. As such, it avoids Mallon and colleagues' criticism.

A more fully developed version of this developmental account of reference promises to be fruitful in different philosophical areas such as: the metaphysics of mind, philosophy of science, and metaethics.[3] Now, exactly how and to what extent this theory proves to be useful in these areas is, of course, an open question.

1.2   Belief and Intentionality

Understanding the nature of thought and belief is necessary to understand how the mind relates to the world and, ultimately, how it is that humans may be said to have knowledge of their surroundings. Much is left to be done in order to achieve such an understanding of the mind, but some progress may be achieved by solving some puzzles of intentionality—i.e., puzzles concerning mental states with content. Some of these longstanding puzzles, owed to Frege 1892 and Kripke 1979, involve proper names.

The evidence presented in chapter 2 suggests a rather simple view of the content of names. According to this view, all there is to the contribution of a name is its referent. Frege 1892 argues that such a view is unable to account for the fact that competent speakers may fully understand the sentence pairs (1a)-(1b) and (2a)-(2b) and still take each sentence to

---

[3]For a more detailed account of how philosophers have drawn important philosophical conclusions from the theory of reference, see Mallon and colleagues 2009.

convey different information even though, unbeknownst to the speaker, (1b) is true.

(1a) Melville is Melville.
(1b) Melville is Bartleby.
(2a) Melville wrote *Moby Dick*.
(2b) Bartleby wrote *Moby Dick*.

If the simple view is correct, then 'Melville' and 'Bartleby' make exactly the same contribution to the content of the sentences in which they are used. How, then, can a competent speaker still take each member of the pair of sentences to convey different information? How can she, for example, believe (1a) but not (1b) or (2a) but not (2b)? It seems obvious that the sentences in question differ in informative value. How can there be such a difference without a corresponding difference in the contribution made by the names?

The substitution of coreferential names fails to preserve *informative* value among simple assertive utterances. The tradition has called this "the problem of informativeness". This problem has given place to further puzzles that are more explicitly concerned with the attitudes. It seems that (3a) and (3b) may be true of the very same subject that ignores (1b)

(3a) Andy believes that Melville wrote *Moby Dick*.
(3b) Andy does not believe that Bartleby wrote *Moby Dick*.

It seems then that the substitution of coreferential names may also fail to preserve *truth* value. Kripke 1979 argues that these problems are more likely concerned with the nature of belief, for a similar puzzle may arise for a competent speaker even when there is a single name involved.

> Peter may learn the name 'Paderewski' with an identification of the person named as a famous pianist. Naturally, having learned this, Peter will assent to "Paderewski has musical talent", and we can infer — using 'Paderewski', as we usually do, to name the Polish musician and statesman:
>
> (3c) Peter believes that Paderewski had musical talent.
>
> Later, in a different circle, Peter learns of someone called 'Paderewski' who was a Polish nationalist leader and Prime Minister. Peter is skeptical of the musical abilities of politicians. He concludes that probably two people . . . were both named 'Paderewski'. [. . .] Peter assents to, "Paderewski had no musical talent". Should we infer . . . :
>
> (3d) Peter believes that Paderewski had no musical talent.
>
> [Kripke, 1979, p.265.]

I deal with these puzzles in Chapters 3 and 4. In chapter 3, I argue that informativeness and substitution failure are not two independent phenomena demanding two distinct expla-

nations. I show they are intimately related, in such a way that any context that gives place to one of them will also give place to the other; and similarly, any solution to any of these problems will also solve the other. I give an account of this intimate relation by looking at research on the cognitive mechanisms involved in what psychologists have called the "theory of mind mechanism" (see Leslie 1987). On this view, there is substitution failure in both language and thought in virtue of it being a design feature of the human cognitive apparatus, a feature that allows our mental states to distinguish between a representation and its interpretation.

In chapter 4, I offer an integrative account of the puzzles. I do so by looking at some evidence on the semantic assumptions that are in play when speakers learn new names and integrating this with a linguistic account of predicate meaning transfer. On the psychological side, I argue that a rational and competent speaker may fail to notice that two distinct expressions are coreferential, even though she may know what both refer to, because she has stored them in different files. I describe some of the criteria that speakers seem to follow when distinguishing among referential expressions. I show how this view is empirically substantiated by offering evidence on memory, information retrieval, and proper name acquisition. An advantage of this view it that it does not make any extraordinary cognitive demands on the side of speakers.

On the linguistic side, the account explains how speakers may use sentences like (2a)-(2b) within belief reports in order to convey the information that the ascribee has failed to notice that two expressions are coreferential. This results in belief reports that use different coreferential names and have different truth-values, such as (3a)-(3b).

(3a) Andy believes that Melville wrote *Moby Dick*.
(3b) Andy does not believe that Bartleby wrote *Moby Dick*.

To explain this I appeal to the mechanism of meaning transfer for predicates *and* to a self-referential use of proper names. On this view, when engaged in the job of ascribing attitudes with substitution failure speakers *use* the relevant names to convey themselves. They do so in a way that hearers may identify that something like this is the right interpretation. In chapter 4 I offer an account of the mechanism by means of which hearers may retrieve such interpretation from the conventional meaning of the terms used.

Frege's puzzles of informativeness and substitution failure have been widely applied. Kalderon 2004, for example, argues that one of the central arguments in contemporary metaethics, namely, Moore's famous "open question argument," is nothing more than a variant of Frege's puzzle. If so, then the integrated account presented in chapter 4 should be of some relevance to metaethics. Others have used the problem of substitution failure to argue against certain views of the metaphysics of mind (see Kim 2006, chapter 4). The

integrated account should also be helpful here.

1.3 Imagination and the use of language

We normally use names to talk about the world of our experience, but we also use them to imagine how that world could be and how we want it to be. We use them not only to engage in acts of pretense (e.g., fictional names) but also in acts of serious science making (e.g., failed scientific hypothesis). Understanding the ordinary use of empty names seems central to understand this rather important part of human cognition. How do speakers manage to use names like 'Santa' and 'Vulcan' to convey truth-evaluable, non-trivial, information?

Children believe in the existence of fictional characters, and behave accordingly. Scientists sometimes endorse failed hypothesis, and behave accordingly. In their relevant contexts, utterances of 'Santa wears red' and 'Vulcan is a planet' may convey something true and informative. How can all these claims be true? What is it that children believe when they believe *that Santa is coming*? What did Le Verrier believe when he thought *that Vulcan was responsible for the changes in the orbit of Mercury*?

To answer these questions we need to explain: (i) what is the contribution made by an empty name to the content of a sentence; (ii) how this makes such content a truth-evaluable one; (iii) how that content is efficacious at guiding behavior; (iv) how this fits in the overall picture of language and cognitive development; (v) how this fits within a general fiction / non-fiction distinction; (vi) and how this fits within a more general semantic theory for proper names. Chapter 5 presents a new account of empty names, the cognitive theory, that is capable of addressing all these issues.

The cognitive theory of empty names comes in two parts. First, it assumes a referentialist semantics for names, according to which the only semantic contribution of a proper name is its referent. In doing so we can extend the same semantic treatment to *all* referential uses of proper names, including empty names. This is a great advantage and, to my mind, an important contribution of my theory to the debate on empty names. A consequence of this homogenous treatment is that assertions, belief reports, and fictional games that make use of empty names are *not* intelligible in virtue of their semantics alone. So we need the second part of the theory to explain how such uses *are* intelligible.

Here I offer a psychological account of the mechanism involved in uses of empty names in a way that is independent of the semantics of the expressions used. On this view, when using empty names speakers assert, believe, report, and desire the contents of their imaginings, which are concerned with what I call "epistemically decoupled units" or EDUs for short.

The theory contributes to the debate on empty names by arguing for two central theses:

Cognitivism: Referential uses of empty names are intelligible not merely in virtue of semantics and / or pragmatics, but mainly in virtue of the cognitive-general resources they recruit (i.e., something like Leslie's 1987 "decoupling mechanism").

Representationism: Two object-directed attitudes, the content of which is cognitively determined (as above), may be directed toward the same object even if the associated representations have been (cognitively) assigned different referents.

In chapter 5, I offer a description of the cognitive mechanism appealed to by the cognitive theory. I explain how it works and how it delivers the relevant representational objects (i.e., EDUs) that we need in order to address issues (i)-(v) above. I also offer a detailed account of this cognitive content by comparing it with the more well known notion of a proposition. On this view, the explanatorily relevant contribution of an empty name is made cognitively not linguistically; this contribution delivers something pretty much like a proposition (i.e., an EDU), which explains both the truth-evaluability of the relevant speech acts and the behavior-guiding feature of the associated mental states. The theory is consistent with the evidence on competence for proper names, presented in chapter 2, as well as with recent psychological research on pretense. The account is also good enough to meet a challenge posed by Walton 1990 with respect to the general fiction / non-fiction divide.

In chapter 5, I show how this machinery helps us understand the ordinary use of fictional names as well as of empty names associated with failed scientific hypothesis. I think the cognitive theory is well suited to illuminate other philosophical puzzles as well. I briefly describe how it can be used to address some problems in metaphysics and metaethics, and how it offers a solution to the longstanding problem of intentional inexistence owed to Brentano 1874.

Finally, chapter 5 concludes by addressing a puzzle raised by uses of empty names in negative existential constructions: i.e., expressions of the form 'X doesn't exist', where 'X' is a referential term. We all know that Santa does not exist. And the same goes for Vulcan. What we know is a piece of non-trivial information about the actual world: it is such that there is no Santa, and no Vulcan. How is it that we manage to convey this information by using names without referents and, *a fortiori*, without truth-conditions?

There are independent reasons to think that negative existential assertions *do not* involve referential uses of empty names. On my view, we are better off understanding negative existentials as involving *self-referential* uses, i.e., where the name, say, 'Santa', is used to refer to itself. Thus, on this view, the speaker conveys some metalinguistic information

about the name used in the negative existential sentence. Strictly speaking, this use of an empty name is not empty: it refers to itself. Yet it is because ordinary *referential* uses of the same name are empty that negative existential assertions are true.

According to this metalinguistic interpretation, an assertion of (4a) in the relevant context will be properly interpreted as (4b).

 (4a) Santa doesn't exist.
 (4b) 'Santa' doesn't name something.

Chapter 5 concludes by describing two mechanisms—meaning transfer and presupposition accommodation—that speakers may exploit in order to get their audience to properly interpret their use of a negative existential assertion.

References

Brentano, F., 1874: *Psychology from an Empirical Standpoint*. London: Routledge and Kegan Paul.

Cumming, S., 2008: "Variabilism". *Phiosophical Review* 117, 4, 525-554.

Frege, G. 1892: "On sense and nominatum". In A. P. Martinich, (ed.) 2001: *The Philosophy of Language*, Oxford: Oxford University Press.

Hall, D.G., 1999: "Semantics and the acquisition of proper names". In R. Jackendoff, P. Bloom, and K. Wynn, (eds) 1999: *Language, Logic, and Concepts*. Cambridge: MIT Press.

Kalderon, M.E., 2004: "Open questions and the manifest image". *Philosophy and Phenomenological Research* 68, 2, 2004.

Kim, J., 2006: *Philosophy of Mind*. Cambridge, MA: Westview Press.

Kripke, S., 1979: "A puzzle about belief". In A. Margalit, (ed) 1979: *Meaning and Use*. Dordrecht: Synthese.

Leslie, S.J., 2007: "Generics and the structure of the mind". *Philosophical Perspectives* 21, 1, 375-403.

Leslie, S.J., 2008: "Generics: Cognition and acquisition". *Philosophical Review* 117, 1, 1-47.

Mallon, R., Machery, E., Nichols, S., and Stich, S. 2009: "Against arguments from reference". *Philosophy and Phenomenological Research* 79, 2, 332-56.

Recanati, F, forthcoming: "Singular thought: In defence of acquaintance". In R. Jeshion (ed.) forthcoming: *New Essays on Singular Thought* Oxford: Oxford University Press.

Reimer, M., 2003: "Reference". In E. Zalta (ed.) *Stanford Encyclopedia of Philosophy*, http:// plato.stanford.edu/ entries/reference

Shatz, M. 1994: *A Toddler's Life*. New York: Oxford University Press.

Shatz, M. 2008: "Language as a consequence and an enabler of the exercise of higher-order relational capabilities: Evidence from toddlers". *Behavioral and Brain Sciences* 31, 145-46.

Walton, K., 1990: *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.

Chapter 2

On Problems with Descriptivism: Psychological Assumptions and Empirical Evidence

This paper offers an empirical assessment of Description Theories of Proper Names.[1] We look at empirical data from various sources, including lexical and cognitive development, memory, and aphasia, to see whether they support Description Theories. It turns out that Description Theories demand much more, in terms of cognitive skill, than what the data suggest. We argue that Description Theories are lacking in empirical support. This undermines their success as philosophical theories for proper names in natural languages.

2.1   Background

There is an ongoing philosophical debate with respect to proper names. The attention is focused on two central questions. First, assuming that proper names refer, how is it that they do? Second, what is the meaning (if any) of a proper name? Theories of names are classified in three main groups, relative to their answers to these questions: Millian, Descriptivist, and Hybrid. In this paper we start with a review of Mill but we will focus on theories of the second kind.[2]

---

[1]This chapter was co-authored with Marilyn Shatz. A version of this chapter is forthcoming in *Mind & Language*.

[2]Hybrid Theories claim that some descriptive information is necessary for reference-fixing purposes but not for meaning-determination purposes. We do not discuss these theories anywhere in this paper.

2.1.1 Mill's view

John Stuart Mill 1864 offered what seemed to be the most intuitive answer to the question of meaning for proper names: names are denotative but not connotative. Mill has been paraphrased as claiming that the only meaning, if any, of a proper name is its referent. This view has been defended and developed by Marcus 1961, Kripke 1980, and Donnellan 1966 and 1970 among many others. (See Reimer, 2003; Braun, 2007; and Lycan, 2006 for a more detailed account.)

Mill's view has one initial flaw. It does not offer a response to the question of reference fixing for proper names. It is not, properly speaking, a complete theory of proper names. This condition has been supplemented by means of what is called a "causal theory of reference" (see Kripke, 1980, and Evans, 1973 for discussion). An important feature of this theory is that it does not require the use of any identifying descriptive information. As a result, the requirements for proper-name use and acquisition are very minimal. All the speaker needs is to stand in the proper causal relation or have the proper referential intentions. The speaker need not identify the descriptive information (if any) that other competent speakers may associate with their use of the name.

2.1.2 Frege's insight

The simplicity of this theory is compelling. Unfortunately, it is also problematic. Frege 1892 describes what seems to be a crucial inadequacy of Mill's view. Consider the following sentences: 'Hesperus is Venus' and 'Venus is Venus'. If they are true, then both express, according to the Millian, exactly the same content: that one object (i.e., the planet Venus) is identical to itself. This is so because true identity statements use coreferential names, and having the same reference is tantamount to having the same meaning.

Intuitively, however, it seems that 'Hesperus is Venus' is more informative, less trivial, than 'Venus is Venus'. Still, they vary only in the use of different but coreferential names. Hence, the difference cannot be explained in terms of reference. Yet, appealing to reference seems to be the only move available to the Millian. After all, there is supposed to be no other feature of names that may be content-relevant.

To solve this problem Frege 1892 proposed a different theory of names. According to this view proper names have reference (Bedeutung) and sense (Sinn). Frege's "sense" is commonly taken as an analogue of a definite description (see Dummett 1973 and 1981) for a somewhat different interpretation). A definite description is an expression of the form 'The F' where 'F' is a predicate satisfied by a unique object. According to Rusell's 1905

traditional account, definite descriptions carry uniqueness and existential entailments. The former entailment is of central importance. It guarantees that the definite description will designate a single individual, thus being an ideal candidate to play the reference-fixing role for proper names. Definite descriptions convey uniquely identifying information about the designated object. An indefinite description, like 'A PLANET', will not do the work. It cannot fix the reference of, say, 'Venus' , because the descriptive information it conveys does not uniquely identify the object it describes.

The resulting theory claims that each proper name has an associated definite description. Definite descriptions are meant to express senses and these are, according to Frege, abstract, external, objects. For the speaker to understand a name she must grasp a sense. According to Frege at least, *what the speaker grasps* determines the reference of her use of the name. This view is usually called "descriptivism". Following Reimer 2003, we call this, and similar theories, "basic descriptivism".

This explains how the reference gets fixed: i.e., the referent will be the object that satisfies the definite description. It does not yet explain what determines the meaning of a name. For that we need Frege's further claim that the reference-fixing mechanism also determines the meaning of the name. To put it somehow, names have two different levels of meaning: the referent and the associated description. Call this, and similar theories, "expanded descriptivism". Since Frege and Russell, a great variety of description theories have been offered. Section 2.2 explains how these theories diverge.

It is important to note that the explanatory power of descriptivism comes at a particular cost. In order to be competent in the use of a (descriptivist) proper name, a speaker must represent or understand the relevant concepts expressed by the appropriate definite description. It seems that speakers can be competent users of proper names with much less resources than those required by descriptivist theories. It is our contention that this claim is confirmed by the experimental data.[3]

Before we proceed, we want to clarify our argument. What does it mean to say that competent speakers know, or understand, something?[4] Following various Description Theorists, we will argue (see section 2.6.1) that Description Theories are concerned with actual

---

[3]Millian theorists —most famously Kripke 1980— have criticized descriptivism on similar grounds. Among other problems, Kripke points at what Devitt and Sterelny (Devitt and Sterelny 1999) call the "ignorance" problem. Briefly put, the claim is that a speaker who ignores exactly which properties identify the object to which their use of a name refers, may nevertheless refer to it successfully.

[4]A traditional answer in philosophy, see Wittgenstein 1953 and Austin 1962, argues that linguistic competence is a matter of language use or knowledge-*how*. Another option is to take competence as an instance of propositional knowledge, knowledge-*that*, which need not involve any actual use. This is still an open debate in philosophy. Generally speaking, linguists subscribe to the competence / performance distinction; thereby opening the door for a use-independent notion of competence. The distinction, however, is not widely accepted among philosophers. See section 2.6.1 for further discussion.

language use and, hence, that their answer to this question should be complemented with empirical research. With this in mind, we use the terms 'competent' and 'competent speaker' simply to refer to a speaker who is able to use a name successfully, for the purposes at hand, and according to the applicable standards.

If so, then Description Theories presuppose, at least, *some* kind of processing model. In section 2.3 we offer a general picture of the cognitive skills required for proper name processing according to Description Theories. These are what we call "the psychological assumptions of descriptivism". Section 2.4 evaluates these assumptions against the empirical evidence.

Table 2.1 Varieties of Description Theories

Description Theories	Rigidified Descriptions	Meaning Determination
Expanded	No: Mental Content	Yes
Cluster	Yes: Cluster	Yes
Boethian	Yes:@-transformed	Yes
Causal	Yes: Causal	No
Two-Dimensional	Yes: Causal and	Partially
	Counterfactual	Two Meanings
Wide Scope	Yes: Mental Content	Yes
Metalinguistic	No	Metalinguistic

(see also footnote 5)

2.2   Varieties of descriptivism

There are seven distinguishable kinds of Description Theories of proper names. See Table 2.1. They all vary depending on how they respond to the central questions: how do names refer? And, what is the meaning of a name?

2.2.1   Expanded descriptivism

Frege 1892 and, under a certain reading, Russell 1905, proposed the view previously described. This view, expanded descriptivism, claims that the mechanism of reference is given by a definite description, and that the meaning of a name is both, the meaning of such description and the referent. Suppose that *S* is a competent user of the name 'N'. This view entails that there is one definite description 'DD' which serves as the reference mechanism

of 'N'. Frege 1892 also claims that for *S* to be competent in the use of 'N', *S* must understand 'DD'.

Thus, expanded descriptivism requires that a particular definite description, 'DD', constitute the reference-fixing mechanism and the meaning of a name; it also claims that a competent speaker must know this.

2.2.2 Cluster descriptivism

This latter requirement proved to be problematic for the reasons above presented (i.e., the problem of ignorance). Subsequent theories have modified their requirements in order to fix this problem. Cluster descriptivism claims that, instead of a single definite description, the reference-fixing mechanism is given by the disjunction of all the definite descriptions that the speaker associates with the use of the name. These might include descriptions such as 'the object that other speakers in my community refer to by 'N' .' In its expanded version, cluster descriptivism claims that this disjunction also defines the meaning of 'N.' Wittgenstein 1953 and Strawson 1959 endorse this view, and Searle 1958 seems to do so (though he has two different stories). The expanded version requires that the speaker take the cluster to be synonymous with 'N' and, hence, understand the disjunctive cluster in order to use 'N.'

Searle 1983 presents a somewhat different account. For this version the reference-fixing mechanism does not constitute the meaning of the name. Thus, basic cluster descriptivism does not require that *S* knows, or is able to single out, such description in order to understand the meaning of 'N.' It does require, however, that *S* be able to identify the referent as satisfying the relevant definite description in order to fix the reference of 'N.' This is the requirement of a basic form of descriptivism. Otherwise, *S* will not be able to use the proposed mechanism in order to fix the reference of 'N.' Furthermore, since expanded descriptivisms include basic descriptivism, the latter requirement is common to both forms of descriptivism.

2.2.3 Boethian expanded descriptivism

Boethian descriptivism claims that proper names express essential properties of their referents and that they do so via descriptions. One way to get this result is through the @-transform of a definite description. The definite description 'the second planet between the Sun and the Earth' uniquely identifies Venus. Now suppose that '@' is a proper name of the actual world. The @-transform of the description above is: 'the second planet between

the Sun and the Earth in @'. This latter description expresses a property that Venus has in all possible worlds (according to Plantinga, an essential property).

Plantinga's Boethian descriptivism (Plantinga, 1978) is also expanded. The reference-fixing mechanism of 'N', say, '@-DD' is the meaning of 'N'. Thus, Boethian descriptivism not only requires that *S* be able to identify the referent as satisfying the relevant @-transformed description, but also that she understands such a description in order to be a competent user of 'N'.

2.2.4   Basic causal descriptivism

Some theorists prefer a basic form of descriptivism without endorsing an expanded version (see Lewis, 1984, Kroon, 1987, and Jackson, 1998a). As presented in Lewis 1984, causal descriptivism claims that the reference-fixing mechanism is constituted by a description that: (1) may be rigidified, as in 'the actual A'; (2) may be egocentric, as in 'the actual A in my town'; (3) may involve mental states, as in 'the actual A in my town that I am thinking of'; and (4) should involve relations of causal acquaintance, as in 'the actual A in my town that is the causal source of my A-thought.'

Call the latter a 'causal, rigidified, description'.[5] Basic causal descriptivism claims that the reference-fixing mechanism of a proper name 'N' is given by a causal, rigidified, description. It does not claim, however, that such a description also determines the meaning of 'N'. Thus, basic causal descriptivism does not require that *S* understands, or identifies, such a description in order to understand the meaning of 'N'. As any other descriptivism, it does require that *S* be able to identify the referent as satisfying the relevant description for reference-fixing purposes.[6] This is how causal description theories differ from Kripke's view. According to the latter, even though the reference-fixing mechanism is causal, the speaker need not be able to identify an object as satisfying any definite description to use a name successfully.

---

[5]A rigidifier is an operator that makes of any given term a rigid designator. A rigid designator is a term that refers to the same object in all the possible worlds in which that object exists —including counterfactual ones— and does not refer in worlds in which it does not exist. The description 'The president of the USA' is not a rigid designator, but the rigidified description 'The ACTUAL President of the USA' is. The latter description, but not the former, refers to George Bush in counterfactual situations where the President is not Bush but Gore.

[6]Nelson 2002 defends what seems to be an expanded causal descriptivism. See Nelson 2002, p.410 on different versions of causal descriptivism. Braddon-Mitchell and Jackson 2007 seem to have a similar story.

2.2.5 Two-dimensional (expanded) descriptivism

Jackson 1998b and Chalmers 2002 defend theories according to which every proper name has two meanings: a uniquely identifying property and the object that instantiates such a property. This version is similar to causal descriptivism in that it does not identify the meaning of a name with the mechanism used to fix its reference. As with causal theories, the description used for reference-fixing may be causal, egocentric, etc. It can also be counterfactual (e.g., 'the object that in such counterfactual situation has such property'). However, the description used to refer to the uniquely identifying property is at least part of the meaning of the name. In this sense, it differs from causal theories by holding a view closer to that of Expanded Description Theories. Two-dimensional theories do not explicitly require that the speaker be knowledgeable about which description partly determines the meaning of the name.

Two-dimensional descriptivism does require that a speaker *S* be *able* to understand the relevant description in order to understand a name 'N' —even though it does not require that *S* knows which description it is. It also requires that *S* be able to identify the referent as satisfying the relevant description.

2.2.6 Wide scope expanded descriptivism

As proposed by Dummett 1981 and, on a certain reading, Russell 1905, this view is consistent with an expanded description theory. It has been developed and strengthened, in different ways, by Sosa 2001 and Stanley 1997. This theory presupposes Russell's 1905 account of descriptions, according to which definite descriptions have two readings when used in opaque contexts. The central claim of wide scope expanded descriptivism is that proper names take a wide scope with respect to modal operators. It is also consistent with the requirement of rigidified descriptions.

As such, wide scope expanded descriptivism has almost the same requirements as expanded descriptivism. *S* must be able to identify the referent as satisfying the relevant description. *S* must also understand the relevant description when using 'N'.

2.2.7 Metalinguistic descriptivism

Metalinguistic descriptivism is meant to be only a theory of the meaning of proper names. As such, it is not committed to any particular theory of the reference-fixing mechanism of names. This makes it a distinct description theory, distinct from both basic and expanded

versions. The central claim of metalinguistic descriptivism is that for any name 'N' there is one single description of the form 'the bearer of 'N' ' such that it is the meaning of 'N'. No other description will do and no other restrictions can be put into the description. The view was initially presented by Russell 1919 and later developed by Loar 1978, and Bach 1981. On this version, metalinguistic descriptivism is only committed to the claim that S must understand 'the bearer of 'N' ' in order to be a competent user of 'N'.

Katz 1994 offers a different version. He accepts the original claim about the meaning of 'N' while adding that such a description is necessary but not sufficient for reference-fixing. On his "purely metalinguistic" view, the sense is given by the description 'the thing which is a bearer of 'N' .' It is clear that this is not sufficient for reference fixing, since there normally are several objects that bear the same name 'N.' [7]

2.3 What descriptivism presupposes

We have presented the diverse claims that each of these versions make. Given the variety of Description Theories it seems difficult to say what exactly they are all committed to. Nevertheless, there is something they are all supposed to have in common. In this section we will try to answer the following questions. First, what (if any) is (are) the common claim(s) underlying all forms of descriptivism? Second, what is required for the claim(s) to be fulfilled?

2.3.1 Towards an operative account of descriptivism

The first question has an easy answer. The different description theories we have presented fall into three different groups: basic, semantic, and expanded. All of them appeal to psychological (or internal) facts to account for referential uses of proper names. Basic versions are committed to REFERENTIAL, semantic versions are committed to SEMANTIC, and expanded versions are committed to both.

REFERENTIAL for any competent subject S, any name *'N'* and some appropriate definite description *'DD'*, S competently understands *'N'* iff S refers to N by means of DD.

SEMANTIC for any competent subject S, any name *'N'* and some appropriate definite description *'DD'*, S competently understands *'N'* iff S understands DD and DD is the

---

[7]What makes of this a purely metalinguistic view is the claim that the property of BEING A BEARER OF 'N' is not supposed to be a property of the object, but simply a relation that holds in virtue of certain linguistic practices.

content of *'N'* for *S*.[8]

The second question, however, is not so easily answered. We want to see if the empirical data support a description theory of names. To do so we need an operative account of description theories. An operative account would give an answer to our second question: what is required to confirm description theories? That, however, is exactly what we are missing. We will offer such an account hoping that it is acceptable to description theorists.

Let us start by stating our assumptions. First, there is something to be said about the theory's goal. Is the theory about some logically possible language or about natural languages? We presuppose it is a theory of the latter kind. Is it a theory of how ordinary competent speakers of any natural language, say, English speakers use proper names? We presuppose a positive answer to this question. If, contrary to what we think, description theories do not purport to describe ordinary uses of proper names, then much of what we have to say in this paper will be irrelevant for such theories. However, there are reasons to think that this assumption is confirmed. For a detailed discussion of this point, see 2.6.1.

If description theories are concerned with ordinary uses of names, then they must presuppose that something like REFERENTIAL and SEMANTIC are true of the way in which ordinary English speakers use proper names. This is our second assumption, which comes with our third and final. We presuppose that description theories should be consistent with the empirical data about ordinary English speakers' use of names, which includes cognitive, lexical, and neurological evidence.

This gives us a place from which to start. An operative account of description theories will, at least, give us the psychological analogues —an operational version— of REFERENTIAL and SEMANTIC. What must *psychologically* be the case in order for these to be true? The rest of this section is devoted to developing the psychological picture of a competent speaker according to description theories of names. In section 2.4 we present three kinds of data (lexical, cognitive, and memory) arguing for a psychology that is incompatible with the characterization presented here. The issue, then, is not whether description theories entail a psychological account; we take that as a given. Rather, the issue is whether they do it in a empirically supported way.

Let us start with what description theorists have to say. According to Stanley:

> If the descriptive picture is true, then, for each expression in our language, we possess a priori uniquely identifying knowledge about its referent. Such a premise is more than just a useful tool in epistemological and metaphysical

---

[8]These two claims together illustrate the classical view of concepts as applied to proper names. See Margolis and Laurence 1999, especially pp.21-23.

theorizing. For if the descriptive picture is true, then we have a rich store of a priori knowledge. [1997, p.565.][9]

Description theories of names require "a rich store' of "uniquely identifying knowledge about [. . .] referents". We think it is fair to rephrase this in the following way. According to description theories, competent speakers have a rich store of *concepts* or *representations*, which they use to identify referents. These concepts are meant to be the cognitive correlate of the central linguistic tool: definite descriptions. (From now on, we will loosely use 'definite descriptions' and, alternatively, 'uniquely identifying concepts' to mean pretty much the same. Nothing central hinges upon the choice of terms.)

There are three different ways of understanding concepts (see Margolis and Laurence, 1999, and Rey, 1994): as mental particulars, abstract entities, or as abilities. Since our central claim is that description theorists demand too much in terms of cognitive skills, we will assume that their view on concepts is the least demanding one; i.e., concepts as abilities. If our claim is true, then description theorists will be in fault whatever their concept ontology may be.

On this view, the a priori knowledge that Stanley talks about turns out to be an ability. Following the Descriptivist's central claim, it is the ability to identify objects by assessing their properties. Notice that this is true for all versions since both, REFERENTIAL and SEMANTIC require the use of such knowledge either to fix reference or to determine meaning.

This gives us our first psychological requirement of descriptivism. To competently understand names is, at least partly, to be able to identify objects by assessing their properties. For example, one way to be competent in the use of 'Venus' is for a speaker to be able to identify Venus as being the causal source of the visual experience that she has when observing the morning star. This is the psychological analogue of REFERENTIAL, which is a commitment of both basic and expanded descriptivism. Let us call this demand, COGNITIVE.

COGNITIVE for any competent subject *S*, any name *'N'*, and the relevant definite description *'DD'*, *S* competently understands *'N'* only if *S* is able to identify *N* as having the unique properties expressed by *'DD'*.

---

[9]Stanley does not explicitly say what he means by 'a priori knowledge'. One way to understand this, a paradigmatic example perhaps, is the kind of knowledge that can be obtained by simply knowing the meaning of a term. This seems to be what Stanley refers to by 'a priori knowledge,' in other words, the knowledge that we have in virtue of the fact that we are competent speakers of a language. What is relevant here is not the account of the *a priori*, but the claim that speakers have a rich store of conceptual knowledge simply in virtue of their linguistic competence.

In section 2.4 we present evidence that is incompatible with this picture. The argument is simple. As we have seen, if description theories are true, then competent users of names must, at least, be able to identify objects by assessing their properties. The evidence in 2.4.1 shows that competent understanding of names starts at about six months of age. However, the data in 2.4.2 shows that property assessment is not in the early repertoire of infants. It follows that competence for names does not require that the subject identifies objects by assessing their properties and, hence, that something is wrong with description theories of names. Section 2.4.3 shows that this separation between object identification and property assessment is not simply eliminated through development. The data here show that even for adults, who have a greater linguistic and conceptual expertise, there is such a split.

There is more to the descriptivist picture. Expanded versions claim that the meaning of a name (i.e., the information it carries) is determined by a definite description and that it *is* in fact the meaning of the latter. In other words, expanded description theories are not only committed to the claim that what is stored (meaning) is the same, but also that the access to it (via descriptions) is the same. This requires a particular information-processing model according to which the meanings of names and descriptions have the same standing with respect to storage and retrieval. In particular, if someone is unable to retrieve the appropriate description —perhaps owed to some peculiar case of aphasia— then she should also be unable to retrieve the relevant proper name; simply because they are, content-wise, the same. At least that seems to be the idea behind Frege's 1892 foundational claim that the cognitive value of names is given by definite descriptions. This gives us another psychological requirement, this time it is the analogue of SEMANTIC. Let us call this principle; MEMORY.

MEMORY for any competent subject *S*, any name *'N'* and some appropriate definite description *'DD'*, *S* cannot competently use *'N'* without retrieving the content of *'DD'*.

In section 2.4.3 we argue against this claim. If expanded description theories are correct then storing and retrieving the content of any given name is tantamount to storing and retrieving that of the relevant definite description. In section 2.4.3 we present evidence of double dissociation (i.e., selective impairment and preservation) of proper names, suggesting that storage and retrieval of names differs from the rest of the lexicon. We also present studies on how non-injured subjects store and retrieve names showing that names are more difficult to recall than common nouns. This suggests that storing and retrieving a name is *not* tantamount to storing and retrieving a definite description and, thus, that their contents are not identical at all. It must be, then, that the claims of expanded description theories are

mistaken.

COGNITIVE and MEMORY give us a psychological picture of basic and expanded description theories of names. See Table 2.2. This picture fails to describe the way in which ordinary speakers use proper names. This undermines their credentials as theories for proper names of natural languages.

Table 2.2   Speakers according to Description Theories

Cognitive	Memory
S is able to identify objects by assessing their properties	*S* cannot retrieve *N* without retrieving *DD*
Basic / Expanded	**Expanded**

2.3.2   A caveat

We want to prevent confusion about metalinguistic descriptivism. There are two versions of it. On Kripke's 1980 version, any given name 'N' is associated with the description 'the referent of 'N' '. Those who defend the theory offer a different description: i.e., 'the bearer of 'N' '. These descriptions denote two different properties. An object can be the referent of a speaker's use of 'N' without thereby being named 'N' (or bearing the name 'N'). Whether an object is a referent of 'N' depends on how a speaker uses 'N'; whether it bears the name 'N' depends on social naming practices. As Geurts puts it, "bearing a name is like bearing a tie. Like ties, names are seldom unique, but circumstances permitting they may be used for referential purposes (. . .). Taken on its own (. . .) a name doesn't refer any more than a tie does" [Geurts, 1997, p.326].

Consequently, a speaker need not know which name has been given to an object (which name it bears) in order to know it is the referent of her use of, say, 'N'. All she needs to do is take the relevant object to be the referent of the name. On Kripke's version, metalinguistic descriptivism is trivially true. What we have to say here does not go against this theory. It does, however, speak against the less trivial version according to which speakers must understand that objects can be given names that they bear.

2.4 The evidence

In this section we review part of the literature on the psychology of proper names. This includes studies in early lexical development, cognitive development, and name retrieval. In so doing we intend to offer a picture of what competent users of proper names look like: which lexical, cognitive and memory skills they use.

2.4.1 Early lexical development

Thanks to studies on early lexical development we know that proper names are among the earliest words in an infant's lexicon and that this is so even though the early lexicon is very limited and generally biased for nouns over verbs (see Nelson, 1973, 1974; Gentner, 1982; Choi and Gopnik, 1993; and Tardif, 2006 for discussion).

Jusczyk and Mandel 1996 studied how young infants (experimental subjects were between four and six months of age) react to familiar sounds, particularly their own names. Various experiments were designed in order to determine how accurately these infants represent their own names. In one experiment the infants were presented with six-sentence passages, one of which included the infant's name and another which included another infant's name with which she had been previously familiarized. The results suggest that six-month-olds "detect the occurrence of their names in fluent speech passages" [Jusczyk and Mandel, 1996, p.36-37]. Jusczyk and Mandel conclude that infants have detailed representations of the sound patterns of their names, and that they detect their own names in fluent speech sooner than they can detect other words. This suggests that names are special lexical items. They do not come hand-in-hand with other words. (See Mandel, Jusczyk, and Pisoni, 1995, for more on this issue.)

This supports our claim that young infants understand names competently. Apparently, there is an important distinction between nouns, especially proper names, and other words. For some or other reason, proper names are easier to learn. Studies on the noun bias show that names do not demand the same cognitive skills as verbs and adjectives.[10] Proper nouns might be easiest of all because they simply require 'N' (or a sequence of sounds) to

---

[10]Gentner 1982, argues that this noun bias is owed to the fact that noun-encoded meanings are easier to learn than verb-encoded ones. Nouns simply refer to things while verbs encode relational-meanings. Gentner argues that there is a natural partition between objects and relations. Following a similar line of argument, Waxman 1992 claims that infants understand that types of words relate to types of meanings, nouns to objects, verbs to relations, and adjectives to properties. Snedeker and Gleitman 2004 claim that it is easier to learn nouns because they require less information from the linguistic context to determine their meaning. Gillette, Gleitman, Gleitman, and Lederer 1999 show that, even for adults, it is easier to make correct guesses about nouns than about verbs.

relate to the bearer of 'N' in no specified way. Nouns require the mapping of sounds onto non-linguistic contexts, which makes them easier to learn than verbs, which require more linguistic information. Proper nouns in particular require the least amount of mapping: the same sounds get mapped into the same unique context; i.e., the bearer. (See Piccin and Waxman, 2007, for a lexical account.)

2.4.2 Cognitive development

Xu and colleagues (see Xu and Carey 1996, and Xu, Carey, and Quint 2004) argue for the psychological independence of object individuation from property assessment in early infancy. They begin by assuming Spelke, Breinlinger, Macomber, and Jacobson 1992 famous results suggesting that an object-concept, i.e., BOUNDED PHYSICAL OBJECT, is genetically encoded. Their goal is to test a hypothesis about the acquisition of descriptive (or sortal) concepts, e.g., DOG, BALL, HAT, which provide criteria for identification and individuation. Owed to Bower 1974, among others, the Object-first Hypothesis claims that infants may use "spatiotemporal criteria for individuating and tracing identity of objects well before they can use other property information" [Xu and Carey, 1996, p.114].

To test their hypothesis, Xu and Carey presented infants (age of experimental subjects was 10 months) with events where an object (e.g., a rabbit) emerged from behind a screen alternating with a second object (e.g., a cup) emerging from the other side of the screen. Infants were presented with two outcomes when removing the screen. The expected outcome presented two different objects, the unexpected one a single object. If infants were able to identify objects by assessing their properties (e.g., shape), they should look longer at the unexpected outcome.

The experiment was carefully designed to include three conditions, each of which ended in either an expected or unexpected result. The first two conditions varied the circumstances of presentation, the third provided a baseline. In the *property* condition, infants were initially presented with one object (e.g., a truck) emerging from the right side of the screen. The object was then returned behind the screen. Infants were then presented with a second object (e.g., an elephant) emerging from the left side of the screen. The second object was returned and the infant was presented with the expected (two objects) or unexpected (one object) outcome. In the *spatiotemporal* condition, the experiment differed in only one feature: infants were simultaneously presented with the two objects (e.g., a truck and an elephant) at the beginning of the experiment. The baseline conditions simply included the presentation of one or two objects coming from behind the screen with no earlier presentation of them.

The baseline conditions showed that infants had an intrinsic preference for (looked

longer at) two objects. Given these conditions, a successful trial required that the infant be able to overcome this preference. The results show that infants did overcome the baseline in the spatiotemporal condition but not in the property one. This suggests that 10 month-olds do have a concept of BOUNDED PHYSICAL OBJECT but no proper understanding of more specific concepts, e.g., properties such as SHAPE.

Xu and Carey 1996 includes a comparative study between 10 and 12 month-olds. The results showed that 12-month olds were able to overcome the baseline preference in the property trial. Xu and Carey hypothesize that this difference is related the child's stage in lexical development. It is between these ages that infants begin learning common nouns.

In a more recent study, Xu 2002 found out that even nine-month-olds could use shape information for object individuation *if* they were given linguistic aid: infants were able to form a representation of two different objects when they were presented with different labels, e.g., 'Look, a duck!' and 'Look, a ball!'. However, in Xu et.al. 2004, Xu and colleagues presented 12-month-olds with a different trial. This time they were presented with two different objects that differed in color but not in shape (e.g., a red and a green ball). The results show that even 12-month-olds were unable to make this distinction. This suggests that at 12-months of age, infants do have a limited understanding of general notions, e.g., PHYSICAL OBJECT, SHAPE, but not enough of them to identify objects by assessing their properties (see Xu et.al., 2004, p. 159). Thus, although it can be argued that 12-month-olds do have at least some general property concepts, the evidence suggests that object individuation and property assessment are two separate psychological processes.

It is important to note, first, that in order to identify an object by assessing its unique properties, infants need an understanding of properties that develops over the second year of life; and, second, that an initial understanding of general properties (12 months) takes place a long time after infants show an understanding of proper names (6 months), as seen in the previous section. Thus, assessing properties is not necessary to fix the referent of a proper name. There must be something wrong with the claims of both basic and expanded versions of description theories.

These claims strongly suggest that infants do not use definite descriptions (or uniquely identifying descriptive concepts) either to fix the referent or to determine the meaning of a proper name. Cognitively speaking, much less is required in learning and using proper names than in uniquely identifying an object by discerning its properties.

2.4.3 Memory for proper names

The suggested separation between reference and property assessment in early cognitive development is confirmed by studies on memory for proper names. Because most of those studies are with adults, they also suggest that this separation is sustained throughout development. Furthermore, they offer an empirical basis from which to argue particularly against semantic and expanded description theories (i.e., those that take definite descriptions to determine the meaning of names).

Adults are *able* to use definite descriptions to fix the reference of a proper name. Kripke 1980 presents 'Neptune' as a case. Some time before astronomers were able to see Neptune, it was hypothesized that something caused the changes in the orbit of Uranus. However, the referent of 'Neptune' could not be determined as any salient unidentified object. The name 'Neptune' was, thus, introduced by means of a definite description, e.g., 'the object that causes the changes in the orbit of Uranus'. There seems to be a difference between adult and infant use of proper names. Is this a qualitative difference? Do adults process names in very different ways? Or, alternatively, is this just a quantitative difference? Do adults simply aid themselves by using more information? An answer to these questions comes from an answer to a different, but related, question: do adults necessarily use definite descriptions in their acquisition and use of proper names? The empirical evidence suggests a negative answer. Adult use of proper names is not qualitatively different from that of infants.

Anderson and Hastie 1974 tested undergraduates on processing information depending on whether it was introduced by a proper name or a definite description. Subjects were given a number of predicates ('F', 'G' and 'H') about a person A and introduced her with a proper name *'N'* such that $N = $ A. They were then given different predicates ('R', 'Q' and 'S') about the same person, but this time the person was introduced by means of a definite description *'DD'*, without the subjects knowing so, such that $DD = $ B. They were then told that A $= $ B. This forced the subjects to make inferences when asked to verify a sentence that used either term (i.e., *'N'* or *'DD'*) followed by the predicates introduced by means of the other term (e.g., having *'N'* followed by 'R', 'Q' and 'S'. Non-inferences (e.g., *'N'* is F') were introduced for comparison.

Inferences with *'DD'* took longer than inferences with *'N'*. In fact, *'N'* inferences took the same time as *'N'* non-inferences. *'DD'* inferences, however, took up to twice as long as *'DD'* non-inferences. Anderson and Hastie claim that this is owed, first, to the fact that when the information is initially stored as belonging to two different individuals, two memory nodes, A and B, are created. Second, they claim that when the identity A $= $ B is established, the proper name *'N'* is directly connected with the second node B; whereas the definite description *'DD'* is only directly connected with the node B and indirectly connected with A

by means of the identity A = B. According to Anderson and Hastie, the inferences went as follows: first, for the proper name there were only two steps from *'N'* to B and from B to the predicates 'R','Q' and 'S'; second, for the definite description, there were four steps, from *'DD'* to B, from B to *'N'*, from *'N'* to A and from A to the predicates 'F','G' and 'H'. This difference in processing suggests that, even for adults, definite descriptions are not necessary for understanding proper names. Further evidence supports this claim.

Names are difficult to remember. Tip-of-the-tongue problems, show this very clearly. Everyone has, at some point, been unable to recall someone's name, while knowing pretty well who that person is. The speaker is acquainted with the referent, she might be able to describe the referent, she also knows the name, and yet the name is not retrieved. Folk-psychology describes this as the tip-of-the-tongue or TOT phenomenon.

The TOT phenomenon has prompted several models on how proper names are processed. Cohen 1990 tested the differences between processing meaningful and meaningless expressions. The results show that recalling proper names is just as difficult as recalling meaningless, non-word, expressions (e.g., 'wesp', 'blick'). Based on this evidence, Cohen argues that the TOT problems with proper names are owed to a lack of semantic associations. On her view, proper names are detached from conceptual representations. Some, see Valentine, Brennen, and Brédart 1996, argue that names are not absolutely meaningless. For example, the name 'Baker' conveys the information that the bearer is of English origin or an English speaker. Even in this case, however, it is generally agreed that proper names are vulnerable in terms of retrieval and that this is owed to the fact that "a proper name conveys almost no information about the entity it names" [Valentine et.al., 1996, p. 108].

Several models intend to account for this evidence in terms of lexical access (see Valentine et.al. 1996 for a detailed survey of these models). Most models require three different kinds of processing units: 1) word recognition units; 2) biographical information, or semantic information, units; and 3) face recognition units. Word recognition units are required to identify the name either from within speech or within the lexicon. Face recognition units are required to identify the object to which the name refers. Semantic information units, however, are not a necessary part of the process. Speakers can in fact have access to such biographical information about the referent of the name, but it is not necessary for the production of proper names. If it were so, then there would be more entry points, within memory, from which to access a name. Proper names would not be difficult to retrieve. In their review of the different models of lexical access for proper names, Valentine and colleagues conclude by defending

> A functional architecture of face naming which is compatible [. . .] with [. . .] the notion that proper names are pure referring expressions. One important fea-

ture of this architecture is that it does not assume that identity-specific semantic information constitutes a necessary pathway to names. Simulations using an interactive activation version of this model produced results consistent with the fact that names are more difficult to retrieve than identity-specific semantic information. [Valentine et.al., 1996, p.139]

The data on the TOT phenomena and the processing models presented suggest that proper names and definite descriptions are independently stored and retrieved. These are good reasons to think that they have different semantic values. More evidence comes from further studies. Cohen and Burke 1993 review the evidence on memory for proper names, suggesting that they are meaningless *and* doubly dissociated from other parts of the lexicon. The latter claim is supported by evidence from cases of anomia and selective preservation of names.

LeDorze and Nespoulos 1989 and Semenza and Zettin 1989 present clinical cases of anomia. The patients suffered brain damage and, as a result, had a dramatic inability to retrieve proper names. Le Dorze and Nespoulos studied 24 aphasics, 5 of which were cases of pure anomia, while Semenza and Zettin present a single patient. Semenza and Zettin's patient was unable to "deal with purely referential non-descriptive semantic relations" (Semenza and Zettin, 1989, p.679). Their patient showed no problems when learning and retrieving other types of words. Semenza and Zettin argue that the best explanation of this is to take names as purely referential expressions: "they denote the individuals or the entities that are called by them, [. . .] but do not describe any property or imply any attribute" (Semenza and Zettin, 1989, p.679). Lucchelli and Renzi 1992 support a similar line of argument. They hypothesize that proper names are tags that permit the identification of their bearers but that, on their own, tell nothing of the properties of these bearers. In a similar way, Hadar, Jones, and Mate-Kole 1987 argue for a disconnection between semantic and phonemic lexicons in cases of anomic aphasia.

Names are not only selectively impaired, they can also be selectively preserved. Van Lancker 1990 and Semenza and Sgaramella 1993 present cases of selective preservation of personal names. In both studies, the subjects were able to recall the names of certain familiar faces, without being able to retrieve any biographical information with respect to the referents. Goodglass and Wingfield 1993 present evidence of a similar phenomenon regarding names of places. Based on this evidence, Semenza and Sgaramella argue that proper names are doubly dissociated from other parts of the lexicon (e.g., common names). Brédart, Brennen, and Valentine 1997 argue that the case for double dissociation is fairly limited to comprehension and, even there, to some cases. However, they do accept (see Valentine, et.al. 1996) that the evidence suggests the existence of preservation of proper

names 'in the context of impaired comprehension of proper names' (Valentine, et.al. 1996 p.81-82). For more recent studies see Hadar, Wenkert-Olenik, and Soroker 1998, for studies on anomia see Kohn, and Goodglass 1985. Kambanaros and Steenbrugge 2006 discuss the differences between nouns and verbs based on anomic and other kinds of aphasia. Avila, Lambon, Parcet, Geffner, and Gonzalez-Darder 2001 study how proper name retrieval may be cued by word similarities; and Diaz, Lindin, Galdo-Alvarez, Facal, and Juncos-Rabadán 2007 offer an ERP study on face recognition and TOT problems in naming.

In sum, studies on memory suggest that uniquely identifying information, the kind of information that definite descriptions carry with them, is not necessary for either storing (acquiring) or retrieving (accessing) a proper name. When it comes to retrieval, proper names are vulnerable in a way in which common nouns are not — consider the problems of anomia, selective preservation, and TOT phenomena. This suggests that proper names are dissociated within semantic memory.

Memory-wise, proper names are expensive while common nouns — and, with them, definite descriptions — are cheap, and this is not because there is a lot of information that is carried by a proper name, as semantic and expanded description theories suggest. The evidence goes in the opposite direction. Names are difficult to retrieve *because they carry so little information*. This is predicted by all the models that offer an account of the TOT phenomena. This seems natural. After all, we do tend to forget names but we do not seem to forget definite descriptions. See Table 2.3.

The data suggest that storing and retrieving proper names is completely different, and independent, from storing and retrieving definite descriptions. This should not be the case if definite descriptions were to determine the meaning, or semantic value, of proper names. If the descriptivist story were true, one should expect the human mind to treat the content of a name in exactly the same way as that of a definite description. After all, the content of any given name is supposed to be one and the same thing as that of an appropriate definite description. This supports our claim that the content of a proper name is not determined by a definite description and that expanded description theories fail on the psychological evidence.

2.5   Preliminary Conclusions

We have argued (see sections 2.2 and 2.3) that description theories — whether basic, semantic, or expanded — are committed to the claim that definite descriptions are *necessary* either for reference fixing (see REFERENTIAL page 17), meaning determination (see

Table 2.3 Actual Memory and Descriptivist Demands for Names

Memory for Proper Names	Description Theories
(1) Name retrieval is difficult: TOT problems, anomia, etc.	(1) Retrieving *N* is just like retrieving *DD*.
(2) Name retrieval does not require access to descriptive information.	(2) Retrieving *N* requires retrieving unique descriptive information.
(3) Names carry very minimal, information about the referent.	(3) *N* carries, unique, descriptive, information.

SEMANTIC page 17), or both. Operational counterparts of these claims — COGNITIVE, and MEMORY— were offered (see section 2.3.1, especially page 19). In section 2.4 we presented empirical evidence on how speakers acquire and use proper names. The evidence suggests that COGNITIVE and MEMORY are unsupported. Description theories lack a proper empirical standing.

We have made a case, on developmental grounds, against COGNITIVE. Proper Names are among the first words of an infant's lexicon. Some minimal understanding of reference is, of course, needed for this to be the case, possibly founded in prelinguistic knowledge or natural preparedness.[11] If uniquely identifying concepts were required for reference-fixing, then an understanding of them should also appear as early in the prelinguistic period (6 months). This however is mistaken. There is evidence of an understanding of properties developing only slowly—let alone uniquely identifying ones developing early. This undermines the empirical standing of description theories. To fix the reference of a name, it is not necessary to identify the object as falling under this or that uniquely identifying concept (i.e., as satisfying the relevant definite description).

We also made a case, on neuropsychological grounds, against MEMORY. Proper names are very costly memory-wise; their acquisition and production does not require access to biographical information about the referent, which is the kind of information that a definite description contributes. Names and descriptions are different for the human mind. Storing

---

[11]Regarding the prelinguistic understanding of reference see Hall 2002. See Wellman and Gelman 1997 and Gelman 1990 on distinctions such as the person/non-person and animate/inanimate. Even psychologists who generally disavow unlearned principles recognize the need for some preparedness. See, for example, Rakison and Lupyan 2008.

and retrieving a name is not tantamount to storing and retrieving a definite description. This further undermines the case of semantic and expanded description theories. Storing and retrieving a name is not tantamount to storing and retrieving a definite description.

2.6   Objections and replies

In sections 2.4 and 2.5 we argued that description theories fail on empirical grounds. In this section we support our claim on different grounds by showing that our criticism is of philosophical relevance. First, we support our assumption stated earlier that philosophical theories of language are concerned with actual language use by showing why this is so. Second, we consider a possible objection from description theories. We give reasons to think that the evidence we have reviewed involves a competent understanding of names. Finally, we deal with the general issue of whether psychological evidence is of any philosophical interest. We briefly mention different areas of philosophical debate where evidence from cognitive psychology has directly changed the course of the discussion. We contend that these arguments show why philosophers should pay attention to the psychological evidence.

2.6.1   Accounting for natural languages

Description theories of names are part of a broader philosophical project concerned with language itself. Lewis 1975 offers a simple formal description: languages are abstract things constituted by a set of ordered pairs of the form <S,M>, where the first object is a sentence and the second a meaning. A satisfactory account of a language's grammar is, then, one that gives us the rules by means of which we get the set of ordered pairs that constitute that language. Speaker competence becomes, on this view, a relation between a person and the set of ordered pairs so construed. It tells us what a person must know in order to understand and produce the relevant ordered pairs. There are infinitely many different languages, natural languages being a subset of them. The latter are, presumably, those languages constituted by the ordered pairs that actual human individuals can learn, understand, and produce. Hence, a satisfactory philosophical account of *natural* language must determine what actual human individuals must know in order to understand and produce the ordered pairs of sentences and meanings that they actually understand and produce. How are we going to determine this? In one way or other we are bound to make reference to actual speakers, those who are competent language users. The question is, what should we look at when referring to actual speakers?

It is important to distinguish two ways in which actual language use can be said to be relevant: *qua explanans* and *qua explanandum*. Judging by the literature, it seems less controversial to claim that language use is relevant in the latter sense. Philosophical theories of natural language describe what is required to use a natural language competently. Description theorists, like Dummett, claim this explicitly—see also Jackson 1998a, p.203; Stanley 1997, p.564; Kroon 2004, p.4-5; and Lewis 1975:

> What we have to give is an account of what a person knows when he knows what a word or expression means, that is, when he understands it. [...] An account of understanding language, i.e., of what it is to know the meanings of words and expressions in the language, is thus at the same time an account of how language functions, that is, not only of how it does what it does, but of what it is that it does. [Dummett, 1973, p.92; see also Dummett, 1993]

We may suppose, then, that actual language use is relevant enough for philosophers to care about *qua explanandum*. The important question is then, whether or not actual language use is relevant *qua explanans*. Is language use a relevant source of explanation to understand what speakers do when they use a language?

Lewis 1992 gives a straight answer that represents the view of other highly influential philosophers such as Stalnaker 1999 and Kaplan 1989 and a standing tradition in philosophy thereof: "Surely it is our use of language that somehow determines meaning" [Lewis, 1992, p.106]. If use determines meaning and linguistic competence is partly, as a matter of fact, knowledge of meaning, then language use becomes a source of explanation of linguistic competence. The arguments in section 2.4 are the result of taking this claim seriously. We agree with Schiffer 2003 and 2006 that a proper account of linguistic competence must include an account of "the information processing that underlies, and thus accounts for, the person's ability to understand utterances in her language" [Schiffer, 2006, p.281.]

So language use turns out to be quite relevant for the philosophy of language. We recognize different ways of studying language use. The studies reviewed in section 2.4 give a good idea of the various approaches one might want to pursue.

2.6.2 A descriptivist reply

It is reasonable to distinguish between what enables a speaker to understand a language and the way in which she actually produces and comprehends her utterances [see Smith 2007, p. 962]. Chomsky's competence / performance distinction has given place to a tradition that

takes evidence of the latter sort to be irrelevant for the former. Can description theorists follow a Chomskian strategy and explain away the counter evidence by taking it to be about performance? We think not.

Recent work in psycholinguistics suggests that competence and performance are not as separate and independent as the quote suggests. "If, for example, a language processor is constrained by limitations of human working memory, that processor itself may have an impact on the very nature of syntactic structure" [Shatz, 2007a, p.6]. (See Lewis and Vasishth 2005 for a case study on the relation between memory and syntax and Diesendruck 2007 for an account of the interactions between the syntactic, semantic, and pragmatic aspects of language acquisition.) It is difficult not to see how the evidence in section 2.4 bears at the level of the function computed.

To illustrate this, let us suppose that description theorists understand names as Chomsky understands grammar. By analogy, then, description theorists can say their theories describe actual speakers' competence, but not actual speakers' performance.

If such were the case, then the resulting arguments would look more or less like this. First, infants in fact know enough about definite descriptions in order to use them when fixing the reference of names; there is no lexical evidence of this because definite descriptions require certain performance skills that infants lack. Second, all speakers in fact understand names as synonymous with some definite description; the evidence goes astray because it measures performance, and performance requires a lot more than just semantic knowledge. In short, the lexical and some of the memory problems here presented would not affect description theories. Such arguments seem to save the day for them. But, do they really?

First, the Chomskian analogy has its limits. It is meant to be about knowledge of language, not about cognitive capacities in general. If such is the case, then description theorists still have to explain how it is that young infants understand *definite descriptions* (competence, not performance) even though they show no evidence of an ability to understand property *concepts and representations*. Claiming that infants have a prelinguistic understanding of that rich store of concepts that definite descriptions require is far-fetched. It goes beyond the claims of even the more rationalist students of cognitive development (see Gelman and Williams 1998 for discussion). Unless description theorists can show that infants know a lot more than what cognitive psychologists tell us about them, it is not clear how they will solve the cognitive problem (see sections 2.4.1 and 2.4.2) simply by claiming to offer theories of competence, not performance.

Second, the problem of memory (see section 2.4.3) would still be standing. Description theories would still have to explain why there is selective impairment and preservation of proper names. Unlike the tip-of-the-tongue phenomena, these do not seem to be problems of

performance. Remember that we are supposing that these theories are, at least, interested in describing *actual* speakers' competence. If for a competent, healthy speaker, proper names and definite descriptions go hand-in-hand, then there should not be selective impairment or preservation in non-normal cases. Any brain damage that affects proper names should also affect definite descriptions and, with them, common nouns. Accounts of anomic aphasia seem to bear upon the proper level of description, that of the function computed.

Perhaps there is a way to circumvent this problem. What if it is not the name, but the route between name and definite description that is damaged? Though less clearly, this might still be viewed as a problem of performance. The speaker cannot produce the name, but still knows it. Would this work?

We do not think so. For that to be the case, names and definite descriptions must be stored in different locales. This, unfortunately, seems inconsistent with semantic and expanded descriptivism, according to which the content of a definite description *is* the meaning of the name. A description theorist should expect that whatever is stored in the name-locale, the access to which is impaired, must be the same as what is stored in the description-locale, the access to which is preserved. If not identity —perhaps that is not the best way to read the claim that a definite description *determines* the meaning of a name— there should be, at least, a necessary connection between the meaning of a name and that of the relevant definite description. Description theories would require an auxiliary hypothesis to explain why, even though they are necessarily related, the brain still benefits from storing names and descriptions separately and can access one but not the other.

2.6.3   If not definite descriptions, what then?

If the descriptivist is correct, competent users of names use definite descriptions. The evidence presents speakers that do not require descriptions to understand proper names. So, either description theories are mistaken or the subjects in question do not understand proper names competently. We need a different account of the use of proper names — i.e., an account of, at least, the reference-fixing mechanism for proper names — that is consistent with the evidence. In this section we offer a sketch of the kind of story we support.

On our view, the problem of reference-fixing can be solved in terms of a natural preparedness of speakers to understand the reference relation. This view is importantly distinct from other non-descriptivist accounts, such as La Palme Reyes, Macnamara, Reyes, and Zolfaghari 1993. First and foremost, unlike La Palme and colleagues, we do not think that common nouns play a *necessary* role. An understanding of common nouns appears much later (12-18 months) than an understanding of names (6 months). Second, we do not

claim that an understanding of reference per se must be innate. We take our claim for a natural preparedness to be easier to support. All that our argument needs is the existence of a prelinguistic ability to understand reference, and a lack of it with respect to definite descriptions. There is evidence for both claims.

On the one hand, there is independent evidence suggesting that reference is unlearned, independent of perceptual characteristics, universal across languages, and primitive for all languages. Most importantly, however, understanding reference seems to require some uniquely human cognitive abilities.

In a recent, and very detailed, review of the literature in comparative psychology, Penn, Holyoak and Povinelli 2008 argue for a distinctively human cognitive capacity: the ability to interpret the environment symbolically, by using higher-order, abstract, relations. If their claim is correct, then this uniquely-human, cognitive endowment, is good enough to account for the natural preparedness of humans to understand and use names. All that speakers need to do is identify a pair of physical objects (e.g., a unique sequence of sounds and an entity) and take one to stand for, or point to the other. This, together with Spelke and colleagues' 1992 finding that a notion of BOUNDED PHYSICAL OBJECT is not acquired through experience, gives us the ingredients we need: i.e., the abilities to both identify objects (referents) and understand abstract relations (reference).

On the other hand, the evidence suggests that understanding properties, and with them, common nouns, is not prelinguistic. Even more, the evidence suggests that proper names and common nouns develop separately, with nouns appearing much later than names. Not surprisingly, there are tracks of this in adulthood, as the evidence from memory on proper names suggests.

We may now face the objection that if description theorists are correct, competent users of proper names must use definite descriptions. The speakers presented all through section 2.4 seem to use names competently without any need of definite descriptions. Then, either description theorists are mistaken or those subjects do not understand proper names competently. There is, however, evidence suggesting that those speakers have everything they need to understand reference. Their human cognitive endowment is such that they do so. Indeed, as the evidence in section 2.4 shows, human infants as young as 6 month-olds demonstrate the integration of the needed components. They are naturally prepared to do so.

2.6.4 Why philosophers should care

Frege 1917 seemed to think that philosophy had nothing, or little, to learn from psychology. Philosophical theories were meant to be logical, not psychological. But things have changed

since. Goldman 1993 describes this shift very well: "while René Descartes and David Hume created their own theories of the mind, contemporary philosophers must give respectful attention to the findings of scientific research" [Goldman, 1993, p.xii].

The recent history is replete with interactions between philosophy and the cognitive sciences. Goldman 1993 and Stich 1994 report many philosophical applications of cognitive science. Philosophy of language is not an exception. Recent work by Leslie 2007 and 2008 shows how important cognitive psychology is for semantics. Other studies have also used empirical data particularly in relation with the theory of reference. See, for example, Segal 2001, Machery, Mallon, Nichols and Stich 2004, Braisby, Franks, and Hampton 1996, Jeshion 2009, and Coates 2009.

Description theories claim that speakers associate a set of uniquely identifying properties with their use of names, either for reference-fixing or for meaning-determination purposes. The empirical findings on actual uses of names show that these cognitive requirements are not a necessary part of the use of names. The evidence clearly shows that description theories of names are mistaken. This is why we think that philosophers should care about the evidence we have reviewed.

2.7   Concluding remarks: Where are we now?

The evidence we have presented shows that description theories are badly equipped as theories of linguistic phenomena. Their claims about the reference and meaning of proper names presuppose more, in terms of cognitive and lexical skill, than what speakers actually require. The studies reviewed give empirical support to our initial claim: a speaker can successfully use a name even if she lacks the cognitive skills that description theorists say she should have.

More recent studies give further support to this claim. Valentine and Darling 2006 present one more case in which names and descriptive information are processed in different ways. Hollis and Valentine 2001 show evidence that there are different kinds of names. Personal and landmark names seem to be purely referencing expressions, while place names seem to give some kind of (perhaps cultural) information. None of them, however, carry uniquely identifying information. Brédart and Valentine 1998 show that names behave differently from terms that do carry descriptive information about the referred object. Brédart and colleagues 2005 confirm the claim that accessing semantic (biographical) information about the referent is not necessary for proper-name processing.

We proposed natural preparedness as an alternative account for proper name acquisition

— hence, as an alternative account of reference-fixing for proper names. Further evidence was given to support this account. Valentine and colleagues 1996 present different models of proper-name processing that are consistent with this view. However, there are still some open questions for both the psychology and philosophy of proper names.

2.7.1 Psychology

On the psychological side of things, we have not said anything as to how this fits into a more general view of language. There is an interesting question with respect to nonhuman animals, as to whether they can understand proper names. It seems uncontroversial, for example, that dogs — and perhaps also chimpanzees — understand their own name. It would be surprising, however, if dogs can understand higher order abstract relations like that of reference. Would this prove that the alternative model offered is mistaken? We believe not, for several reasons.

First, there are good reasons to think that language use is a uniquely human ability. Shatz 1994 and 2007b presents evidence that even toddlers use language in a way that shows an understanding of higher order, relational, abstract reasoning, and even Theory of Mind abilities such as understanding false belief. Penn and colleagues' 2008 forceful argument supports this view. They offer a great array of evidence on behalf of the distinctiveness of the human mind. Their view encompasses different areas of human cognition including language, Theory of Mind, logic, and causal thinking. Humans are able to "reinterpret the world in a symbolic-relational fashion" by understanding higher-order, abstract relations. Their central claim is that this cognitive capacity, "evolved in only one lineage - ours. Nonhuman animals didn't (and still don't) get it" [Penn, et.al., 2008, p.129].

Second, the very same feature that makes language a uniquely human ability gives plausibility to the preparedness account. Understanding symbols and reference requires an understanding of higher-order abstract relations —e.g., understanding that the word 'Jon' refers to Jon. Others have argued that more advanced uses of proper names (e.g., identity statements) require an understanding of higher-order relations as well (Perner, Rendl and Garnham, 2007). Thus, even if infants look like dogs in their early understanding of names, given that they, and only they, progress beyond that at least raises the question of whether there are no underlying differences between infants and dogs in name understanding from the very beginning. This possibility leads to another question: what do nonhuman animals understand when they identify their names?

Penn and colleagues 2008 have an answer here. According to them, even though there is functional discontinuity between humans and nonhuman animals, there is still no unex-

plainable gap in between. They claim that "different species, as well as different modules within the cognitive architecture of a given species, approximate different features of a PSS [Physical Symbol System] to varying degrees" [Penn, et.al. 2008, p. 128]. It may be that nonhuman use of names is importantly different by not requiring the use of higher-order relations, but yet approximate enough to be regarded as some kind of comprehension of names (see Shatz, 2008).

2.7.2 Philosophy

On the philosophical side of things, we have not said anything with respect to other elements from the debate on proper names. These include modal contexts, belief reports, fictional names, and negative existential statements, among others. Anti-descriptivists like Kripke 1980 and Soames 2002 have argued that names and definite descriptions work differently within modal contexts. For example the name 'Aristotle' and the description 'the last great philosopher of antiquity' behave differently within the sentential context 'X is necessarily the last great philosopher of antiquity'. Psychological studies on counterfactual reasoning might shed some light on how it is that names are used in these contexts (see, for example Perner, et.al., 2007).[12]

Description theorists have argued that the way names behave within belief reports (e.g., 'Aristotle believed that Plato was fond of dogs') supports their view. Others, like Soames 2002 and Stalnaker 1999, argue that this behavior should not be explained in terms of the semantics of proper names. The evidence on semantic memory for proper names seems to support something like the latter view. This, of course, requires further discussion. Something similar can be said about the use of fictional names (e.g., 'Pegasus') and negative existential statements (e.g., 'Plato does not exist'). These are cases where the names used lack a referent. How is it, then, that speakers manage to use them meaningfully? Studies on pretense and fictional reasoning, as well as specific studies on the use of fictional names, may prove to be of help here —Everett and Hofweber 2000 offers a good introduction to the philosophical debate on these issues.

This paper has one further important limitation. It makes a contribution to the philosophical debate about theories of reference-fixing for proper names. This, however, is just one among various non-descriptivist theories (see, for example, Devitt, 1981, and Evans, 1982). We do not offer an adequate discussion of such theories. Doing so is certainly desirable. Unfortunately, it is also beyond the scope of this paper.

---

[12]There is, of course, philosophical literature on this topic. Unfortunately it is too extended to be mentioned here. For an introduction, see Reimer 2003 and Lycan 2006.

2.7.3 Conclusions

Our claim is still standing. Description theories are ill-equipped as theories of what competent speakers understand when using proper names. We have offered considerable evidence against the claim that learning how to use a proper name requires an understanding of uniquely identifying properties, or that names convey uniquely identifying information about their referents. Description theories fail both as theories of reference-fixing and theories of meaning for proper names in natural languages.

References

Anderson, J., and Hastie, R. 1974: "Individuation and reference in memory: proper names and definite descriptions". *Cognitive Psychology* 6, 495-514.

Austin, J. L. 1962: *How to Do Things with Words*. Oxford: Oxford University Press.

Avila, C., Lambon Ralph, M.A., Parcet, M.A., Geffner, D., and Gonzalez-Darder, J.M. 2001: "Implicit word cues facilitate impaired naming performance: evidence from a case of anomia". *Brain and Language* 79, 2, 185-200.

Bach, K. 1981: "What's in a name". *Australasian Journal of Philosophy* 59, 371-186.

Bower, T.G. 1974: *Development in infancy*. San Francisco: W.H. Freeman.

Braddon-Mitchell, D., and Jackson, F., 2007: *Philosophy of Mind and Cognition. An Introduction*. Blackwell Publishing.

Braisby, N., Franks, B., Hampton, J., 1996: "Essentialism, word use, and concepts". *Cognition* 59, 247-74.

Braun, D. 2007: "Names and natural kind terms." In E. Lepore and B.C. Smith (eds), 2007: *Handbook of Philosophy of Language*. Oxford: Oxford University Press.

Brédart, S., Brennen, T., and Valentine, T., 1997: "Dissociations between the processing of proper names and common names". *Cognitive Neuropsychology* 14, 2, 209-17.

Brédart, S., and Valentine, T., 1998: "Descriptiveness and proper name retrieval". *Memory* 6, 2, 199-206.

Brédart, S., Brennen T., Delchambre, M., McNeill, A., Burton, A.M., 2005: "Naming very familiar people: when retrieving names is faster than retrieving semantic biographical information". *British Journal of Psychology* 96, 2, 205-14.

Chalmers, D. 2002: "On sense and intension". *Philosophical Perspectives* 16, 135-182.

Chomsky, N. 1966: *Cartesian Linguistics*. New York: Harper Row.

Choi, S., and Gopnik, A., 1993: "Nouns are not always learned before verbs: an early verb spurt in Korean." In E.V. Clark, (ed.) *The Proceedings of the 25th annual Child Language Research Forum*. New York: Cambridge University Press.

Coates, R., 2009: "A strictly Millian approach to the definition of the proper name". *Mind & Language* 24, 4, 435-46.

Cohen, G. 1990: "Why is it difficult to put names to faces". *British Journal of Psychology* 81, 287-97.

Cohen, G., and Burke, D.M. 1993: "Memory for proper names: a review". *Memory* 1, 4, 249-63.

Devitt, M. 1981: *Designation*. New York: Columbia University Press.

Devitt, M., and Sterelny, K. 1999: *Language and Reality*. Cambridge: MIT Press.

Diaz, F., Lindin, M., Galdo-Alvarez, S., Facal, D., Juncos-Rabadán, O. 2007: "An event-related potentials study of face identification and naming: the tip-of-the-tongue state". *Psychophysiology* 44, 50-68.

Diesendruck, G., 2007: "Mechanisms of word learning." In E. Hoff and M.S. Shatz (eds.) 2007: *Handbook of Language Development*. London: Blackwell.

Donnellan, K. 1966: "Reference and definite descriptions." *The Philosophical Review* 75, 281-304.

Donnellan, K. 1970: "Proper names and identifying descriptions". *Synthese* 12, 335-358.

Dummett, M. 1973: *Frege Philosophy of Language*. New York: Harper and Row.

Dummett, M. 1981: *The Interpretation of Frege's Philosophy*. London: Duckworth.

Dummett, M. 1993: "What do I know when I know a language?" In M. Dummett, 1993: *Seas of Language*. Oxford: Oxford University Press.

Evans, G. 1973: "The causal theory of names". *Proceedings of the Aristotelian Society. Supplementary Volume* 47, 187-208.

Evans, G. 1982: *The Varieties of Reference*. Oxford: Oxford University Press.

Everett, A., and Hofweber, T. (eds.) 2000: *Empty Names, Fiction, and the Puzzles of Non-existence*. Stanford: CSLI Publications.

Everett, A. 2005: "Recent defenses of descriptivism". *Mind & Language* 20, 1, 103-139.

Frege, G. 1892: "On sense and nominatum." In A. P. Martinich, (ed.) 2001: *The Philosophy of Language*, Oxford: Oxford University Press.

Frege, G. 1917: "The thought." In Beaney, M. (ed.) 1997: *The Frege Reader*. Oxford: Blackwell.

Gelman, R. 1990: "First principles organize attention to and learning about relevant data: number and the animate-inanimate distinction as examples". *Cognitive Science* 14, 79-106.

Gelman, R. and Williams, E.M. 1998: "Enabling constraints for cognitive development and learning: Domain specificity and epigenesis." In D. Kuhn and R.S. Siegler, (eds) 1998: *Cognition Perception and Language, Vol. 2* W. Damon (ed.) *Handbook of Child Psychology* 5th edition. New York: Wiley.

Gentner, D. 1982: "Why nouns are learned before verbs: linguistic relativity versus natural partitioning". In S.A. Kuczaj, (ed.) 1982: *Language Development: Syntax and Semantics*. New Jersey: Hillsdale.

Gillette, J., Gleitman, H., Gleitman, L., Lederer, A. 1999: "Human simulations of vocabulary learning". *Cognition* 73, 135-76.

Goldman, A.I. 1993: *Philosophical Applications of Cognitive Science*. USA: Westview Press.

Goodglass, H., and Wingfield, A. 1993: "Selective preservation of a lexical category in aphasia: Dissociations in comprehension of body parts and geographical place names following focal brain lesion". *Memory* 1, 4, 313-28.

Hadar, U., Jones, C., and Mate-Kole, C. 1987: "The disconnection in anomic aphasia between semantic and phonological lexicons". *Cortex* 23, 3, 505-17.

Hadar, U., Wenkert-Olenik, D., and Soroker, N. 1998: "Gesture and the processing of speech: neuropsychological evidence". *Brain and Language* 62, 1, 107-126.

Hall, D.G. 2002: "Semantics and the acquisition of proper names" In R. Jackendoff, P. Bloom, and K. Wynn, (eds.) 2002: *Language, Logic and Concepts*. Cambridge: MIT Press.

Hollis, J., and Valentine, T. 2001: "Proper-name processing: Are proper names pure referencing expressions?". *Journal of Experimental Psychology* 27, 1, 99-116.

Jackson, F. 1998a: "Reference and description revisited". *Philosophical Perspectives* 12, 201-218.

Jackson, F. 1998b: *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. New York: Oxford University Press.

Jeshion, R., 2009: "The significance of names". *Mind & Language* 24, 4, 372-405.

Jusczyk, P.W., and Mandel-Emer, D. 1996: "What's in a name?: how infants respond to some familiar sound patterns". http:// hincapie.psych.purdue.edu /Jusczyk/pdf/Name.pdf.

Kambanaros, M. and van Steenbrugge, W. 2006: "Noun and verb processing in Greek-English bilingual individuals with anomic aphasia and the effect of instrumentality and verb-noun name relation". *Brain and Language* 97, 2, 162-177.

Kaplan, D., 1989: "Demonstratives" in J. Almong, H. Wettstein, and J. Perry (eds.) 1989: *Themes from Kaplan*. New York: Oxford University Press.

Katz, J. 1994: "Names without bearers". *Philosophical Review* 103, 1-39.

Kohn, S.E., and Goodglass, H. 1985: "Picture-naming in aphasia". *Brain and Language* 24, 2, 266-283.

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

Kroon, F. 1987: "Causal descriptivism". *Australasian Journal of Philosophy* 65, 1-17.

Kroon, F. 2004: "Descriptivism, pretense, and the Frege-Russell problems". *Philosophical Review* 113, 1, 1-30.

La Palme Reyes, M., Macnamara, J., Reyes, G.E., and Zolfaghari, H. 1993: "Proper names and how they are learned". *Memory* 1, 4, 433-55.

Le Dorze, G., and Nespoulous, J. 1989: "Anomia in moderate aphasia: problems in accessing the lexical representation". *Brain and Language* 37, 3, 381-400.

Leslie, S.J., 2007: "Generics and the structure of the mind". *Philosophical Perspectives* 21, 1, 375-403.

Leslie, S.J., 2008: "Generics: cognition and acquisition". *Philosophical Review* 117, 1, 1-47.

Lewis, D. 1975: "Languages and language." In D. Lewis 1983: *Philosophical Papers, Vol. 1*. USA: Oxford University Press.

Lewis, D. 1984: "Putnam's paradox". *Australasian Journal of Philosophy* 62, 1-17.

Lewis, D. 1992: "Meaning without use: reply to Hawthorne". *Australasian Journal of Philosophy* 70, 1, 106-10.

Lewis, R., and Vasishth, S. 2005: "An activation-based model of sentence processing as skilled memory retrieval". *Cognitive Science* 29, 375-419.

Loar, B. 1978: "The semantics of singular terms". *Philosophical Studies* 30, 353-77.

Lucchelli, F. and Renzi, E. 1992: "Proper name anomia". *Cortex* 28, 221-230.

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

Machery, E., Mallon, R., Nichols, S., and Stich, S., 2004: "Semantics cross-cultural style". *Cognition* 92, B1-B12.

Mandel, D.R., Jusczyk, P.W., and Pisoni, D.B. 1995: "Infants' recognition of the sound patterns of their own names". *Psychological Science* 6, 315-18.

Marcus, R.B. 1961: "Modalities and intensional languages". *Synthese* 13, 4, 303-322.

Margolis, E., and Laurence, S. 1999: "Concepts and cognitive science." In E. Margolis and S. Laurence (eds.) 1999: *Concepts. Core Readings*. Cambridge, MA: MIT Press.

Matthews, R. J. 2003: "Does linguistic competence require knowledge of language." In A. Barber (ed.) 2003: *Epistemology of Language*. Oxford: Oxford University Press.

Matthews, R.J. 2006: "Knowledge of language and linguistic competence". *Philosophical Issues, 16* 200-20.

Mill, J.S. 1864: *A System of Logic*. New York: Harper and Brothers.

Nelson, K. 1973: "Structure and strategy in learning to talk". *Monographs of the Society for Research in Child Development* 38, 1/2, 1-135.

Nelson, K. 1974: "Concept, word, and sentence: Interrelations in acquisition and development". *Psychological Review* 81, 267-85.

Nelson, M. 2002: "Descriptivism defended". *Nous* 36, 3, 408-36.

Penn, D.C., Holyoak, K.J., and Povinelli, D.J. 2008: "Darwin's mistake: explaining the discontinuity between human and nonhuman minds". *Behavioral and Brain Sciences* 31, 109-78.

Perner, J., Rendl, B., and Garnham, A. 2007: "Objects of desire, thought, and reality: Problems of anchoring discourse referents in development". *Mind and Language* 22, 5, 475-513.

Piccin, T. B. and Waxman, S.R. 2007: "Why nouns trump verbs in word learning: New evidence from children and adults in the human simulation paradigm". *Language, Learning and Development* 3, 4, 295-323.

Plantinga, A. 1978: "The Boethian compromise". *American Philosophical Quarterly* 15, 129-38.

Rakison, D.H., and Lupyan, G. 2008: *Developing object concepts in infancy: an associative learning perspective. Monographs of the Society for Research in Child Development* 73, 1.

Reimer, M., 2003: "Reference". *Stanford Encyclopedia of Philosophy*, http:// plato.stanford.edu/ entries/reference

Rey, G. 1994: "Concepts". In S. Guttenplan (ed.) 1994: *A Companion to the Philosophy of Mind*. Oxford-Cambridge: Blackwell.

Russell, B. 1905: "On denoting." In A.P. Martinich (ed.) 2001: *The Philosophy of Language*. Oxford: Oxford University Press.

Russell, B. 1919: *Introduction to Mathematical Philosophy*. London: George Allen and Unwin.

Schiffer, S. 2003: "Knowledge of meaning." In A. Barber (ed.) 2003: *Epistemology of Language*. Oxford: Oxford University Press.

Schiffer, S. 2006: "Two perspectives on knowledge of language". *Philosophical Issues 16* 275-87.

Searle, J., 1958: "Proper names". *Mind* 112, 447-92.

Searle, J., 1983: *Intentionality* Cambridge: Cambridge University Press.

Segal, G., 2001: "Two theories of names". *Mind & Language* 16, 5, 547-63.

Semenza, C., and Sgaramella, T.M. 1993: "Production of proper names: A clinical study of the effects of phonemic cueing". *Memory* 1, 4, 265-80.

Semenza, C., and Zettin, M. 1989: "Evidence from aphasia for the role of proper names as pure referring expressions". *Nature* 342, 678-79.

Shatz, M. 1994: *A Toddler's Life*. New York: Oxford University Press.

Shatz, M. 2007a: "On the development of the field of language development". In E., Hoff, and M., Shatz, (eds.), 2007: *Handbook of Language Development*. London: Blackwell.

Shatz, M. 2007b: "Revisiting *A toddler's life* for *The Toddler Years*: Conversational participation as a tool for learning across knowledge domains." In C. A. Brownell, and C. B. Kopp (eds.): *Socioemotional development in the toddler years*. New York: Guilford.

Shatz, M. 2008: "Language as a consequence and an enabler of the exercise of higher-order relational capabilities: Evidence from toddlers". *Behavioral and Brain Sciences* 31, 145-46.

Smith, B.C. 2007: "What I know when I know a language." In E. Lepore, and B.C. Smith (eds.), 2007: *The Oxford Handbook in the Philosophy of Language*. Oxford: Oxford University Press.

Snedeker, J., and Gleitman, L. 2004: "Why is it hard to label our concepts?" In G. Hall and S. Waxman (eds.), 2004: *Weaving a Lexicon*. Cambridge: MIT press.

Soames, S. 2002: *Beyond Rigidity*. Oxford: Oxford University Press.

Sosa, D. 2001: "Rigidity in the scope of Russell's theory". *Nous* 35 1-38.

Spelke, E.S., Breinlinger K., Macomber, J., and Jacobson K. 1992: "Origins of knowledge". *Psychological Review* 19, 4, 603-32.

Stalnaker, R. 1999: *Context and Content*. Oxford: Oxford University Press.

Stanley, J. 1997: "Names and rigid designation." In R. Hale and C. Wright (eds.) 1997: *A Companion to the Philosophy of Language*. Oxford: Blackwell.

Stich, S. 1994: "Psychology and Philosophy." In S. Guttenplan (ed.) 1994: *A Companion to the Philosophy of Mind*. Oxford-Cambridge: Blackwell.

Strawson, P.F. 1959: *Individuals*. London: Methuen.

Tardif, T. 2006: "But are they really verbs? Chinese words for action." In K. Hirsh-Pasek and R.M. Golinkoff (eds.) 2006: *When Action Meets Word: How Children Learn Verbs*. New York: Oxford University Press.

Valentine, T., Brennen, T., and Brédart, S. 1996: *The Cognitive Psychology of Proper Names*. London: Routledge.

Valentine, T., and Darling, S. 2006: "Competitor effects in naming objects and famous faces". *European Journal of Cognitive Psychology* 18, 5, 686-707.

Van Lancker, D. 1990: "Preserved recognition of familiar personal names in global aphasia". *Brain and Language* 39, 511-29.

Waxman, S. 1992: "The development of an appreciation of specific linkages between linguistic and conceptual organization". *Lingua* 92, 229-57.

Wellman, H.M. and Gelman, S.A., 1997: "Knowledge acquisition in foundational domains." In D. Kuhn and R. S. Siegler (eds.) 1997: *Cognition Perception and Language. Handbook of Child Psychology, 2*. New York: Wiley.

Wittgenstein, L., 1953: *Philosophical Investigations*. London: MacMillan.

Xu,F. 2002: "The role of language in acquiring object kind concepts in infancy". *Cognition* 85, 223-50.

Xu, F. and Carey, S. 1996: "Infants' metaphysics: The case of numerical identity". *Cognitive Psychology* 30, 111-53.

Xu, F., Carey S., and Quint, N. 2004: "The emergence of kind-based object individuation in infancy". *Cognitive Psychology* 49, 155-90.

Chapter 3

The Puzzle of Names

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

Introduction

Frege 1892 presents the problem of informativeness as a problem about some uses of identity statements. He does not present it as a problem of belief ascription or propositional attitudes. Russell 1905 presents the problem of substitution failure as a problem about belief ascription or propositional attitudes. He does not present it as a problem of informativeness. This way of posing the problems has proven fruitful. Since Frege and Russell, many—with the probable exception of McKay and Nelson 2005 and Swanson 2006—have followed them in conceiving *informativeness* and *substitution failure* as two separate and independent issues; thus, allowing us to understand better what these problems are about.

    This is evident in the way in which different philosophers have tried to solve the problems. Millians seem to agree that they must deal with two different issues. Kripke 1979 addresses the problem of substitution failure, but does not say much (if anything) about informativeness. Something similar goes on in Soames 1989 and 2002. Likewise with descriptivists. In his presentation of what he dubs "the Frege-Russell problems" Kroon 2004 takes them to be different, independent problems. Even contemporary reviews of the topic (see Braun 2007 and Lycan 2006) still present them as two separate issues.

    I believe we have learned enough. Informativeness and substitution failure are not two separate, independent, issues. They go hand-in-hand. They are so closely related to each other that one wonders whether they are really two different problems. In section 3.1 I

will present each problem with some detail. Section 3.2 offers four different solutions from philosophers with opposing theoretical views. The evidence suggests that the alleged "problems" are intimately related since *any* solution to one problem is proved to be enough to solve the other. Section 3.3 gives a brief description of the historical context. It seems that philosophers, in general, have failed to notice this intimate relation. In section 3.4 I offer an explanation, in psychological terms, of why these issues are so intimately related. I conclude in section 3.5 by describing this relation and reformulating the problems in a way that points toward the kind of solution required.[1]

3.1 The puzzles

3.1.1 Informativeness

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

(1a) Melville is Melville.
(1b) Melville is Bartleby.
(2a) Melville wrote *Moby Dick*.
(2b) Bartleby wrote *Moby Dick*.

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

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

---

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

from expanding our knowledge. We are left with the unclear idea of "knowledge expansion".

In a more recent rendition of the problem, Braun 2007 says that (1a) and (1b)—and *a fortiori* (2a) and (2b)—differ in informativeness because "a rational, competent, speaker could understand both and yet think that one is true and the other false" [p.492]. Notice, however, that the criterion of informativeness is psychological: two sentences differ in informativeness if a normal subject can give opposing credences to both (e.g., believing one and disbelieving the other). Thus, mental states play a central role in this understanding of informativeness. Indeed, this is what most philosophers do. They either explicitly appeal to belief and other mental states, as Wettstein 1989 does, or presuppose them, as Kroon 2004 and Lycan 2006 do.

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

(3a) Andy believes that Melville wrote *Moby Dick*.
(3b) Andy does not believe that Bartleby wrote *Moby Dick*.

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

INF: For any pair of sentences S^1 and S^2, which are not necessary falsehoods and which vary only in the use of different coreferential names, a normal subject A, and a context c, S^1 and S^2 differ in informativeness in c iff A believes one and does not believe the other in c.

This account has a limitation. There are many ways in which an utterance of mine can be informative. For example, it can be informative in virtue of its presuppositions or by updating the information my audience has about my mental states (e.g., when you already know who won the prize and I truly say: Jon won the prize). The problem of informativeness, and my account thereof, is not concerned with any of these. It focuses on the way in which an utterance can be informative in virtue of its *subject matter*. A good feature of **INF** above is that it explains not only the difference of informativeness among necessary truths, like (1a) and (1b), but also among contingent truths, like (2a) and (2b).

Two important claims follow from what I have said. First, informativeness is not a property that sentences have *simpliciter* but relative to a speaker—if you prefer we can say it is a relation.[2] This claim is of great importance. If it is true, then differences in informativeness are directly tied to attitudinal mental states.

---

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

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

(2a) Melville wrote *Moby Dick*.
(2b) Bartleby wrote *Moby Dick*.
(3a) Andy believes that Melville wrote *Moby Dick*.
(3b) Andy does not believe that Bartleby wrote *Moby Dick*.
(4a) Andy believes that Melville is Melville.
(4b) Andy does not believe that Melville is Bartleby.

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

3.1.2 Substitution failure

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

(5a) Andy believes that Melville wrote *Moby Dick*.
(5b) Andy does not believe that Bartleby wrote *Moby Dick*.
(5c) Melville is Bartleby.

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

One can choose among many options to solve the problem. Two obvious ones are: (i) to claim that names contribute more than just their referent, or (ii) to directly reject the

possibility of substitution failure (e.g., claiming that (5a) and (5b) cannot both be true. Descriptivists, like Frege, have followed (i) claiming that the semantic contribution of a name is not exhausted by its referent but includes a description. Millians, who believe that names semantically contribute just their referent, have followed both (i) and (ii): the former with non-semantic strategies (see Stalnaker, 1987, Recanati, 1993, and Swanson, 2006) and the latter with different counterintuitive moves (see Braun, 2006, Salmon, 1986 and 2006, Soames, 2002, and plausibly Marcus, 1981). One of the counterintuitive results of the latter strategy is that normal subjects turn out to be either illogical or irrational.[3]

Following option (i) requires one to accept that subjects may ascribe incompatible properties to the same object. If the sentences in (5) are true, it seems safe to infer that Andy ascribes both properties AUTHORING *Moby Dick* and NOT AUTHORING *Moby Dick* to the same object, namely, Melville. Accepting this is certainly not the same as ascribing irrationality to normal subjects, but it is still problematic, specially if, as is generally admitted, normal subjects with normal conceptual abilities do it. So we have a further problem. Substitution failure is caused by lousy property attributions: what explains this phenomenon? The traditional answer is to claim, explicitly or not, that normal subjects may sometimes fail to notice the truth of some identity statements. This seems especially true of cases like (1a)-(1b) and (4a)-(4b), where it is obvious that competent speakers do ignore such information.

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

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

---

[3]Salmon 2006 and Braun 2006 seem to follow this path. They distinguish between illogical and irrational attitudes. Surprisingly enough Salmon does not say what illogical amounts to, see Salmon 1989 and 2006. Braun 2006 claims that a rational speaker may accept the identity statement (5c) and the substitution of names in (5a) and (5b). Doing so is rational. This entails a logical contradiction, however. Braun accepts this is illogical. He does not say what the difference is between being "illogical" and being "irrational." Schiffer 2006 and 1987 offers objections to Braun and Salmon.

substitution failure among the following lines.

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

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

Second: parallel to the case of informativeness, the failure to substitute *salva veritate* concerns only the subject matter of the belief report: i.e., the ascribee's mental states. Substitution failure is not something that concerns the embedded sentence as it may appear outside of a belief report. Presumably, there is no substitution failure in such cases (e.g., 'Melville is Melville' and 'Melville is Bartleby' have the same truth-value).

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

(4a) Andy believes that Melville is Melville.
(4b) Andy does not believe that Melville is Bartleby.
(5a) Andy believes that Melville wrote *Moby Dick*.
(5b) Andy does not believe that Bartleby wrote *Moby Dick*.
(5c) Melville is Bartleby.

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

In this and the previous section I have shown how differences in informativeness come with substitution failure and vice versa. This suggests that, in so far as they are problematic,

informativeness and *substitution failure* go together. In section 3.2, I will present four different solutions to the problems. I will describe how they are meant to solve one or the other and show how they also solve the remaining problem, which they were not (at least explicitly) meant to solve. This suggests that not only the problems, but also their solutions, go hand-in-hand.

3.2 Case studies

3.2.1 Frege

Frege 1892 explicitly solves *informativeness* by appealing to "senses." He does not explicitly offer a solution to *substitution failure*, but his views on indirect speech offer an explanation. My goal will be to show that any one of these moves, either appealing to indirect speech or to "senses," is good enough to solve both problems.

Frege on informativeness

Frege 1892 claims that there is only one way to account for differences in informativeness:

> A difference could arise only if the difference of the signs corresponds to a difference in the way in which the designated objects are given. [Frege, 1892, p.199]

And later on

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

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

(1a) Melville is Melville.
(1b) Melville is Bartleby.

If sentence (1b) is true, then 'Melville' and 'Bartleby' have the same referent. Yet they differ in informativeness. How can this be? Answer: very simply, the names have different senses that determine the sentences' meanings. 'Melville' presents Melville as, say, *the F*;

whereas 'Bartleby' presents Melville as, say, *the G*. Sentence (1a) conveys the information that *the F is the F*, which is trivial. Sentence (1b) conveys the information that *the F is the G*, which is informative. End of story.[4]

Now, let me ask, does this account for substitution failure? Remember our other problematic sentences:

(5a) Andy believes that Melville wrote *Moby Dick*.
(5b) Andy does not believe that Bartleby wrote *Moby Dick*.
(5c) Melville is Bartleby.

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

This much is clear by now: Frege's account of the differences in informativeness *also* explains why there is substitution failure. This, however, is not something Frege claims explicitly. He seems to have a separate treatment of substitution failure.

Frege on substitution failure

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

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

The embedded sentences in (5a)-(5b) are used indirectly. Frege's theory says that the expressions used in those sentences, say, 'Melville' and 'Bartleby,' change their reference. They now refer to their customary senses, say, *the F* and *the G* respectively. How can we use this to account for substitution failure? Answer: very simply, the names are used indirectly, so, according to Frege, they refer to their customary senses. They have different customary

---

[4]I do not intend to take Frege's story as a version of descriptivism. I simply use definite descriptions in pretty much the same way as Frege does: as a way to represent a *sense*.

senses. Hence, they refer to different things. So the principle of substitutivity does not apply. Problem solved.[5]

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

Suppose all we have is Frege's theory of indirect speech. It says that names with different senses in direct speech will not be coreferential within belief reports. This offers a simple solution to the problem of informativeness. Why do (1a) and (1b), or (2a) and (2b) differ in informativeness? Answer: we know the names 'Melville' and 'Bartleby' are coreferential in all of them. That is why (1a)-(1b) and (2a)-(2b) have the same truth-value. Nevertheless, the indirect speech theory tells us that they are not coreferential in (5a)-(5b), because they cannot be substituted *salva veritate*. Thus, says the indirect speech theory, the names must have different senses. If so, then 'Melville' and 'Bartleby' do not convey the same information. This explains why (1a)-(1b) and (2a)-(2b) differ in informativeness. Problem solved!

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

---

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

3.2.2 Recanati on belief reports

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

Recanati 1993 agrees that names *semantically* contribute only their referent, but claims that referring expressions, in general, *pragmatically* contribute a way of presenting their referent. As with Frege's "senses," it is unclear what these "modes of presentation" are. Recanati suggests that 'Melville,' for example, presents Herman Melville as 'MELVILLE' or, alternatively, as CALLED 'MELVILLE'. It is this pragmatically conveyed information which help "enrich" what is otherwise the underdetermined content of the embedded sentence. The mode of presentation, whatever that is, is added to the content of the embedded sentence. This is, claims Recanati 1993, a pragmatic process that becomes truth-conditionally relevant.

This is how the account explains substitution failure:

> Suppose someone says (5a). If the hearer contextually assumes that the ascriber is being 'faithful' to the believer, he will be led to assume that the ascribed belief concerning Melville involves a mode of presentation [. . .] which includes the descriptive concept 'called "Melville".' Hence, the hearer will assume that the believer thinks of Melville as 'Melville;' he will take the reference of the 'that'- clause as a quasi-singular proposition involving a certain (*de re*) concept over and beyond the reference of [. . .] 'Melville.' [Recanati, 1993, p. 358]

And what is a "quasi-singular" proposition? Assuming propositions are structured sets of objects (e.g., referents and properties), the proposition expressed by 'Melville wrote *Moby Dick*' within (5a) is something like (5ϕ):

(5a) Andy believes that Melville wrote *Moby Dick*.
(5ϕ) $<<$Melville, 'Melville'$>$, AUTHORING *Moby Dick*$>$

Similarly, the proposition expressed by 'Bartleby wrote *Moby Dick*' in (5b) is something like (5ψ):

(5b) Andy does not believe that Bartleby wrote *Moby Dick*.
(5ψ) $<<$Melville, 'Bartleby'$>$, AUTHORING *Moby Dick*$>$

If (5ϕ) and (5ψ), respectively, are the contents of the embedded sentences in (5a) and (5b), then substitution failure should be expected: using different names, 'Melville' or 'Bartleby', delivers different contents and, hence, belief reports that may have different truth-values.

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

(1a) Melville is Melville.
(1b) Melville is Bartleby.
(2a) Melville wrote *Moby Dick*.
(2b) Bartleby wrote *Moby Dick*.

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

Even though both sentences of the pair express the same proposition, (1a)-(1b) and (2a)-(2b) do so by pragmatically conveying *different* modes of presentation of Melville. While (2a), for example, presents the author of *Moby Dick* as 'Melville,' (2b) presents the same object as 'Bartleby.' Hence, the difference in informativeness between both sentences. Similar considerations apply to (1a)-(1b) and their difference in informativeness.

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

3.2.3 Plantinga and the Boethian compromise

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

Plantinga 1978 claims of proper names: (a) that they express properties; (b) that these

properties are essential to the name's referent; and (c) that "different proper names of an object can express logically equivalent but epistemically inequivalent essences of that object." [p.130]. (a) is meant to capture Frege's intuitions while (b) satisfies the Boethian compromise.[6] As for (c), things are more complex.

Plantinga relies on a simple view of essential properties: P is an essential property of *x* iff in every world where *x* exists *x* has P *and* in no world is there an object *y*, distinct from *x*, which has *P*. According to this view, any given object has a plethora of essential properties. One can easily transform a contingent property of a given object into an essential one by turning them into "world-indexed" properties. In this sense *being the author of "The Boethian compromise" in the actual world* would be an essential property of Plantinga.

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

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

(1a) Melville is Melville.
(1b) Melville is Bartleby.
(2a) Melville wrote *Moby Dick*.
(2b) Bartleby wrote *Moby Dick*.

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

Plantinga's solution is simple: accept (c). In other words, we must accept that, even though they express logically equivalent contents, 'Melville' and 'Bartleby' *may* express epistemically inequivalent contents. Plantinga does not say much about what makes two properties be "epistemically inequivalent". He does have a very simple, and very Fregean, criterion for making distinctions of epistemic equivalence between properties.

As you may recall, someone might believe (2a) without believing (2b). Andy is, in fact, such a person. According to Plantinga, Andy's different propositional attitudes towards (2a) and (2b) are sufficient to show that (2a) and (2b) express different propositions: "and

---

[6]According to Plantinga, Boethius was the first philosopher to claim "that names express individual essences" see 1978, p.128. He calls this view "Boethianism".

this is due to the fact that their singular terms express different properties." 'Melville' and 'Bartleby' both express essences of Melville, "but epistemically inequivalent and hence different essences." [Plantinga, 1978, p. 129-30].

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

(5a) Andy believes that Melville wrote *Moby Dick*.
(5b) Andy does not believe that Bartleby wrote *Moby Dick*.

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

Why is there substitution failure? Answer: very simply, proper names express essential properties of objects. Coreferential names may express epistemically inequivalent properties. When such is the case, sentences that differ merely in using different coreferential names may express different propositions. When this happens, there is substitution failure.[7] End of story.

Plantinga also offers an account of informativeness. In his own words,[8]

> Since 'Melville' and 'Bartleby' express epistemically inequivalent essences of Melville, (1a) and (1b) express epistemically inequivalent propositions, so that (1b) can be informative.

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

As you can see, Plantinga's solution to the problem of informativeness is the same as Plantinga's solution to the problem of substitution failure. To make this even more clear, note Plantinga's own remarks concerning his account of informativeness: "Surely this is the natural and intuitively plausible position; surely a person could believe (1a) and (2a)

---

[7]This solution, however, is somehow circular. There is substitution failure because sentences using different coreferential names may express different propositions. This begs the question: how do we know they express different propositions? Plantinga's answer is, now, circular: because someone may believe what is said when using one name without believing what is said when using the other name (i.e., because substitution fails).

[8]Plantinga, 1978, p. 134. I have modified the text to fit the examples of this paper. In the original text Plantinga has 'Hesperus' for 'Melville' and 'Phosphorus' for 'Bartleby'. Sentences (1a) and (1b) correspond to his (26) and (19), respectively.

without believing either (1b) or (2b)."[9] If I am not misreading Plantinga, he claims that "the natural and intuitively plausible position" about differences in informativeness is the one that accounts for substitution failure within attitudinal contexts.

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

3.2.4   Stalnaker on two-dimensionalism

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

$$
\begin{array}{c|cc}
A & w_1 & w_2 \\
w_1 & T & T \\
w_2 & F & F
\end{array}
$$

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

Three pragmatic rules describe the way in which this goal is achieved: the proposition asserted must be true according to some, and false according to other, possibilities in the context set;[10] a proposition must be expressed and have a truth-value relative to all possibilities in the context set; and the same proposition must be expressed relative to each possibility within the context set. It might be that the original context defines a matrix with necessary propositions (i.e., violating the first rule) and different ones relative to different worlds (i.e., violating the third rule). That is the case of matrix *A*. It gives us several candidates

---

[9]Plantinga, 1978, p. 134. Again, I have modified the original text to fit the examples.

[10]These "possibilities" are usually understood as "possible worlds" which represent possible ways for our world to be. They are not, in this sense, individual or partial possibilities of this or that thing being thus or so. They represent, rather, the totality of the world as being thus or so.

to the content of the assertion. We cannot pick a proposition because our direct rules of interpretation do not yet tell us what was asserted.

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

$$
\begin{array}{c|cc}
†A & w_1 & w_2 \\
\hline
w_1 & T & F \\
w_2 & T & F
\end{array}
$$

Matrix $†A$ advises us to exclude possibility w_2 and keep w_1, thereby reducing the context set in a single, unambiguous, way. Stalnaker 1978 argues that this model is good enough to explain the differences in informativeness. Consider our problematic sentences (1a) and (1b).

(1a) Melville is Melville.
(1b) Melville is Bartleby.

Suppose I say (1b) to Andy, "Melville is Bartleby". Now suppose there are two different possibilities as part of the context: w_1 where Melville is Bartleby and w_2 where 'Melville' and 'Bartleby' refer to two different people who happen to be close friends. This generates matrix B with two propositions, both of which are trivial.

$$
\begin{array}{c|cc}
B & w_1 & w_2 \\
\hline
w_1 & T & T \\
w_2 & F & F
\end{array}
$$

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

> Now if we try to bring the initial context set into conformity with the third principle by shrinking it, say by throwing out world [w_2], we will bring it into conflict with the first principle by making the assertion trivial. But if we look at what is actually going on in the example, if we ask what possible states of affairs the speaker would be trying to exclude from the context set if he made that statement, we can work backward to the proposition expressed. A moments reflection shows that what the speaker is saying is that the actual world is [w_1] and not [w_2]. What he means to communicate is that the diagonal proposition of the [original] matrix [...] is true. [Stalnaker 1978, p. 91]

This explains how (1b) is informative and (1a) trivial and, thus, how these two sentences differ in informativeness. Assertions determine two-dimensional matrices that must be

compatible with the speaker's presuppositions. Asserting (1b) generates a matrix with necessarily true and necessarily false propositions, thus giving place to a contingent diagonal. This diagonal proposition is informative. Asserting (1a) generates a matrix with only necessarily true propositions, thus giving place to a necessarily true diagonal. Neither this nor the horizontal propositions of the matrix are informative. That is why (1a) and (1b) differ in informativeness. End of story.

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

(5b) Andy does not believe that Bartleby wrote *Moby Dick*.

Suppose that w_1 is the actual world, w_2 is a world in which 'Melville' and 'Bartleby' refer to different people (none of which is the actual Melville) and where Melville wrote *Moby Dick*, and w_3 is similar to w_2 but here Bartleby wrote *Moby Dick*. This gives us matrix C for the embedded sentence 'Bartleby wrote *Moby Dick*' in (5b).

C	w_1	w_2	w_3
w_1	T	F	F
w_2	F	F	T
w_3	F	F	T

Now consider the case of (5a), which determines matrix D for the embedded sentence 'Melville wrote *Moby Dick*'.

(5a) Andy believes that Melville wrote *Moby Dick*.

D	w_1	w_2	w_3
w_1	T	F	F
w_2	F	T	F
w_3	F	T	F

Given this distribution over possibilities, we are asked to reinterpret both assertions as saying something true. As a result, (5b) asks us to exclude possibility w_2 and keep possibilities w_1 and w_3, as illustrated by the diagonal of matrix C, whereas (5a) asks us to exclude possibility w_3 and keep possibilities w_1 and w_2, as illustrated by the diagonal of matrix D.

This accounts for substitution failure by explaining why we cannot substitute 'Melville' for 'Bartleby' even if they are correferential in the actual world. Assertions determine two-dimensional matrices that must be compatible with the speaker's presuppositions. The embedded sentence in (5b) conveys the proposition illustrated by the diagonal of matrix **C**.

The embedded sentence in (5a) conveys the one illustrated by the diagonal of matrix **D**. The matrices differ, that is why the names are not substitutable. End of story.

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

This, however, is (if at all) a problem for the two-dimensional proposal of Stalnaker. It is certainly not the problem we are interested in. Suppose, for the sake of the argument, that we find a way to solve this problem i.e., a way to define a matrix compatible with both the speaker's and the subject's presuppositions.[11] What would be the solution to the problem of substitution failure?

Suppose I utter (5a) and (5b), with the corresponding substitution failure among the embedded sentences. When uttered in the actual world, they contradict each other. What does Stalnaker's theory say about them? Define a matrix compatible with Andy's presupposition that *Melville is not the same person as Bartleby*. Include counterfactual possibilities, like w_2 and w_3 above. The resulting matrices represent different propositions expressed in different worlds. Diagonalization yields different derived matrices for the different embedded sentences. They do not contradict each other, they simply express different, logically consistent, propositions.

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

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

---

[11]Stalnaker does seem to have a rather obvious way to do this: "The propositional concept we construct is one not for the sentence as [the ascribee] would use or understand it, but for the sentence as the speaker and addressee would use and understand it if they were in the possible worlds relative to which the propositional concept is being defined." [1987, p. 127].

3.3 A historical account

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

Frege 1892, for example, seems to be a paradigmatic case of those who take informativeness and substitution failure to be two separate issues. As I have shown, his two solutions are completely different, even though each could have been generalized to account for the remaining case. The same seems to happen with Recanati 1993. He offers a solution to substitution failure and says close to nothing about informativeness. Given that he is an advocate of direct reference, it would be strange for him to do this if he had thought the issues were so closely related. Similar considerations apply to Barwise and Perry 1983, Crimmins and Perry 1989, and Salmon 1986. Soames' 2002 case is different. His account of informativeness is closely related to his account of substitution failure. But the way he presents them, one of them being a problem posed by propositional attitudes but not the other, suggests that he takes them to be two different problems. The same goes for Lycan's 2006 and Braun's 2007 recent introductions to the philosophy of proper names. Both agree that Millianism faces four different challenges, informativeness and substitution failure being two of them. They do not say if the problems are related, and certainly do not seem to think they are *closely* related at all.

Things are clearer with Russell 1905. He knows Frege's informativeness puzzle well enough. Nonetheless, he does not consider it among the puzzles that any good theory should solve; all he deals with is substitution failure.[12] According to Russell, Frege's puzzle is simply about "why it is often worthwhile to assert identity." [1905 p. 214]. This suggests that Russell assumed that they were two different problems. The same goes on with Quine 1980. His goal is to deal with the problem of substitution failure, which he considers to be owed to the presence of referential opacity. There is no reason to think that Quine is unaware of Frege's informativeness puzzle, yet he says nothing about it, even though identity statements appear all over the text.

With Kripke 1979 things are a bit unclear. He presents a new puzzle, this time about

---

[12]It is important to say that I am *not* claiming, or implying, either that Russell did not know there was some such puzzle or that he did not have a way of explaining it. It may very well be that Russell does not pay attention to this puzzle, say, because he thinks it does not arise for idealized languages. If such is the case, then better for me. If this historical view of Russell is correct then Russell did think that they were two different unrelated problems. My point is not that Russell fails to account for the puzzle. Not at all. My point is that Russell fails to notice that these two puzzles are so intimately related with respect to *natural* languages that it is hard to see why they are two distinct puzzles.

belief, involving substitution failure, but does not offer any solution to any of, what I take to be, closely related problems. He seems to be concerned pretty much only with substitution failure, though it could be argued that he is assuming that informativeness is closely related to it.

Finally, there are other philosophers, like Plantinga 1978, Stalnaker 1978 and 1987, and Jacob 2003, who seem to think that the problems are closely related, somehow. Plantinga 1978 does think that a proper solution should account for both, but he clearly takes them to be two different problems that cannot be solved by non-Fregeans. Stalnaker 1978 and 1987 offers pretty much the same solution to both problems, but he explicitly says that these are two separate issues, given that substitution failure presents one further obstacle having to do with the ascription of attitudes. Jacob 2003 presents both issues as problems of intentionality. He takes them to be related to each other but says nothing about their relation, and it is fair to say that he takes them to be different issues.

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

3.4   Why is there substitution failure?

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

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

(1a) Melville is Melville.
(1b) Melville is Bartleby.
(2a) Melville wrote *Moby Dick*.
(2b) Bartleby wrote *Moby Dick*.

What does this imply? Recall our formulation of informativeness:

INF: For any pair of sentences S^1 and S^2, which are not necessary falsehoods and which vary only in the use of different correferential names, a normal subject A, and a context c, S^1 and S^2 differ in informativeness in c iff A believes one and does not believe the other in c.

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

Now, suppose that, for some unexplained reason, among the inhabitants of world w_δ there is a linguistic practice according to which all coreferential terms are substitutable *salva veritate* within attitudinal contexts. It is important to note that this is *only* a change in linguistic practice and is meant *not* to include any cognitive or psychological difference between human beings here in the actual world and the inhabitants of w_δ. In particular, suppose that speakers in this world use the term 'schmelieves' that we define as:

schmelieve: For any subject A and any names *'m'* and *'n'*; if *m=n* then A schmelieves that *m is F* iff A schmelieves that *n is F*.

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

(4a) Andy schmelieves that Melville is Melville.
(4c) Melville is Bartleby.
(4d) Andy schmelieves that Melville is Bartleby.

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

Ascribing schmeliefs to subjects may very well help illuminate their behavior. If all you know about *Moby Dick* is that Bartleby wrote it, and you wonder why Andy keeps buying different editions of *Moby Dick* I can explain his behavior by saying: *Andy schmelieves that Bartleby is the best author of the English tradition*. Yet, ascribing schmeliefs is, by far, not *as* explanatory as ascribing beliefs. It cannot explain why, for example, normal

agents assent to one but not the other of the problematic pair of sentences (1a)-(1b). Andy's distinctive assents to the distinct sentences is not an inference from the ascriber's use of 'schmelieves;' it is a puzzling fact that cannot be explained unless 'schmelieves' allows for some hyperintentionality, which it precludes by definition.

To further illustrate this point, consider the following extension of the scenario above. Andy loves *Moby Dick* and he wants to meet the author. One day he is told that Bartleby will be presenting a book in the coffee shop downstairs. He assents to both 'Melville wrote *Moby Dick*' and 'Melville is Melville', but not to 'Melville is Bartleby'. Given this, the definition of 'schmelieves' above, and the fact that Melville=Bartleby, we should conclude that Andy schmelieves that Melville will be presenting a book in the coffee shop downstairs. Now, suppose further on, that Andy assents to 'Bartleby is the worst writer of the English tradition.' Given this situation, it is terribly unclear what we can do (if anything at all) to understand and predict Andy's behavior on the basis of schmelief ascription. Should we expect him to go to the coffee shop downstairs? After all, Andy schmelieves that Melville is the best writer of the English tradition and he schmelieves that he will be presenting a book downstairs. Yet, Andy was told that Bartleby would be the one presenting and he does assent to 'Bartleby is the worst writer of the English tradition.' Should we, then, expect him not to go downstairs? We have, I am afraid, no clue of what to expect from Andy.

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

Leslie (see Leslie 1987, 1994, and 2002) notes that one of our characteristically human cognitive abilities, that of pretending that something is the case, depends upon the ability to understand "object substitution" : i.e., the ability to assign props, or stand-ins, as the referents of representations. Object substitution requires that humans be able to raise a representation, say and create an uninterpreted copy of it. Leslie 1987 takes this uninterpreted copy to be the mental analogue of a quoted sentence.

To illustrate, consider a case of pretense involving object substitution: to pretend of a banana that it is a telephone. To do so, one must be able to go from the representation *that this is a telephone* to representing *that 'this is a telephone'*. Once we have the latter,

different props may be assigned as referents of, say, 'this': e.g., a banana. At this point one can pretend *that this is a telephone*, even when 'this' *refers to* a banana. It would be a mistake, however, to interpret 'this' as *contributing* the banana to the interpretation of the representation *this is a telephone*. Interpreting the representation this way would give place to mistaken inferences: e.g., *that this telephone is a fruit*. Mental states exhibiting object substitution, like this one, preclude such reference-based interpretations and inferences, since referents are precisely what is substitutable in them.

Leslie 1987 forcefully argues that object substitution is closely related with the referential opacity of "mental state terms such as *believe*, *expect* and *want*" [Leslie 1987, p.416]. This "referential opacity" is nothing more than the phenomenon of substitution failure.

> For example, "the prime minister of Britain" and "Mrs. Thatcher" refer at the time of this writing to the same person. Therefore, anything asserted about the prime minister of Britain, if true, must be true of Mrs. Thatcher as well. If it is true that the prime minister of Britain lives at No.10 Downing Street, then it must be true that Mrs. Thatcher lives at No.10 Downing Street. But put this proposition in the context of a mental state term and this no longer holds. Thus, "Sarah-Jane believes that the prime minister of Britain lives at No.10 Downing Street" in no way entails the truth (or falsehood) of "Sarah-Jane believes that Mrs. Thatcher lives at No.10 Downing Street." In a mental state context one can no longer "look through" terms to see what they refer to in deciding such issues. The mental state term suspends normal reference relations." [Leslie 1987, p.416]

The relation between substitution failure in mental state reports and object substitution in pretense is clear: in both cases it would be wrong to interpret the relevant referential terms as contributing their referents; in both cases, reference-based interpretations would yield mistaken inferences. Unlike with 'schmelieve', there is an "isomorphism" between hyperintensional mental state terms such as 'believe' and mental representational states such as pretense. The former exhibit substitution failure, the latter exhibit object substitution. Leslie 1987 claims this isomorphism is "not coincidental". To explain it Leslie posits

> An underlying form of internal representation that possesses these semantic properties [i.e., referential opacity]. Mental state expressions and pretense both depend cognitively on these representations and therefore inherit their properties. [Leslie 1987, p.416]

This "underlying form of internal representation" has been shown to be relevant not only for pretense (see Leslie 1987 and Friedman and Leslie 2007), but also for the development of an understanding of mental states (see Baron-Cohen, Leslie, and Frith 1985) and, with it, for the development of moral reasoning (see Wellman and Miller 2008).

I think this gives us a very good idea of what would be missing if we were to have just the cognitive correlate of non-hyperintensional attitude verbs such as 'schmelieve.' If cognitive psychologists are correct, there would be no understanding of pretense, false belief, and, apparently also, social-practical reasoning.

Consider, for example, what would be of pretense if all we were to have were schmeliefs. Suppose Jon plays the role of *Oedipus*, Carla plays that of *Jocasta*, and Michael plays *Laius* in a representation of Sophocles' drama. To understand the representation I must come to schmelieve *that Oedipus killed Laius and married Jocasta*. But, as a matter of fact, Jon is Edipus, Carla is Jocasta, and Michael is Laius. Assuming that all there is to mental representational states are schmeliefs, it follows that to understand the representation I must also schmelieve that Jon killed Michael and married Carla. But this is, very clearly, wrong. If all we had were non-hyperintensional mental states like schmeliefs, we would not be able to understand acts of fiction making.

If Leslie 1987, 2002, and Friedman and Leslie 2007 are correct, then "mental hyperintensionality" turns out to be a design feature of human cognition. If so, and assuming attitudinal verbs are part of a broader machinery for explaining human cognition and predicting human behavior, it is not surprising that a good number of attitudinal verbs are hyperintensional. These, I hope, are good enough reasons to think we do need something else than just schmeliefs and schmelief-ascription. A world with differences in informativeness but no substitution failure seems to require some substantial departure from what seems to be actual human psychology.

3.5   A working hypothesis

The ability to have mental states exhibiting object substitution sheds some light over our problematic phenomena: informativeness and substitution failure. As we saw with pretense, these mental states are such that the proper interpretation of the representation involved is not fixed by the semantically determined reference. Suppose that a single actress plays all the central roles in Sophocles' drama. In such a case I can understand *that Oedipus killed Laius* without believing (or pretending) *that Laius killed Oedipus*, even if I know that 'Oedipus' and 'Laius' refer to the same actress. Furthermore, I must do so if I am to understand the theatrical representation. It is exactly this kind of mental state that is involved in cases where there are differences in informativeness and failures of substitution.

As it is clear from the example above, mental states involving object substitution allow for "reference" substitution, but not for "sign" substitution. In the theatrical representation,

that Oedipus killed Laius does not entail *that Laius killed Oedipus* even if, in that context, 'Oedipus' and 'Laius' refer to the same individual (i.e., a virtuous actress). A subject would fail to understand the theatrical representation if she made this kind of inference. These are contexts where a difference in signs (e.g., between 'Oedipus' and 'Laius') matters for the interpretation even without a corresponding difference in reference. This difference in signs can amount to a difference in informativeness or give place to failures of substitution.

If our representational mental states were to map *only* into the semantically determined interpretation, differences in informativeness and failures of substitution would be a mystery. We would have to complicate the semantics in order to account for them. But we have independent reasons to avoid following this line, at least with respect to proper names (see García-Ramírez and Shatz forthcoming).[13]

If, however, we allow representational mental states to map into something else than just the semantically determined interpretation, then informativeness and substitution failure are a natural result. Two or more sentences may not differ in their semantic content and yet differ with respect to the signs they use to convey that content. And it may very well be that this difference of signs is the "something else" that representational mental states map into.

Recall our formulation of the problem:

INF: For any pair of sentences S^1 and S^2, which are not necessary falsehoods and which vary only in the use of different coreferential names, a normal subject A, and a context c, S^1 and S^2 differ in informativeness in c iff A believes one and does not believe the other in c.

The claim was: if a normal speaker assents to (or believes) one sentence without assenting to (or believing) the other, we have reasons to think there are differences in informativeness. If we take the relevant representational mental states to map into something else (e.g., the signs themselves) than just semantic content, we would have an explanation of why there are different attitudes (e.g., assent and dissent) without there being semantic differences. All we need is for the representations to involve different, yet coreferential, signs.

The same goes for substitution failure:

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

---

[13]It is worth noting that some traditional ways to complicate the semantics (i.e., by appealing to descriptive information) would not even solve the problem. Consider the case of the single person enacting Sophocles' drama. Even if we take 'Oedipus' to mean "the F" and 'Laius' to mean "the G", a competent subject would still be allowed to go from believing (or pretending) *that Oedipus killed Laius* to believing (or pretending) *that Laius killed Oedipus*, since she would know that, as a matter of fact, the F = the G.

The claim was: two coreferential names fail to substitute within attitudinal contexts whenever the relevant subject fails to accept the truth of the corresponding identity statement. Once more, if we allow for representational mental states to map into something else than just the semantically determined interpretation we get an explanation of substitution failure.

If there are, for example, belief-like mental states that map not only into semantic differences but also into sign differences, then the latter may be relevant to belief formation: e.g., a speaker may believe *that N is N* and disbelieve *that N is M* even though 'N' and 'M' are coreferential. Now, if different (yet coreferential) signs guide the speaker's belief formation in different ways, then different (yet coreferential) signs will *not* be interchangeable within reports of the speaker's belief states. Reporting the belief states of such a speaker in the relevant context will involve substitution failure.

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

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

This formulation prompts many questions, all of which I shall leave unanswered here. Which strings of sounds count as different names? How can a representational mental state map into something else than the semantically determined interpretation? What needs to be the case, cognitively speaking, for there to be differences of informativeness and substitution failure relative to a normal subject *A*? All of these should be addressed by a substantial account of the phenomena in question. I do not intend to do such thing in this paper, but in its sequel (see García-Ramírez forthcoming). I will be satisfied to have shown that the apparently distinct problematic phenomena of informativeness and substitution failure are intimately related and why this is so.

References

Baron-Cohen, S., Leslie, A.M., and Frith, U., 1985: "Does the autistic child have a "theory of mind"?" *Cognition* 21, 37-46

Barwise, J., and Perry, J., 1983: *Situations and Attitudes*. Cambridge: MIT press.

Bealer, G. 1993: "A solution to Frege's puzzle". *Philosophical Perspectives* No. 7, *Language and Logic* 17-60.

Braun, D. 2006: "Illogical but rational". *Nous* 40, 2, 376-9.

Braun, D., 2007: "Names and natural kind terms." In E. Lepore and B.C. Smith (eds.), 2007: *Handbook of Philosophy of Language*. Oxford: Oxford University Press.

Crimmins, M., and Perry, J., 1989: "The prince and the phone booth". *Journal of Philosophy* Vol.86, No.12, p.685-711.

Frege, G., 1879:"Begriffsschrift." In M. Beaney, (ed.), 1997: *The Frege Reader*. Oxford: Blackwell.

Frege, G., 1892: "On sense and nominatum." In A.P. Martinich, (ed.), 2002: *The Philosophy of Language*. Oxford: Oxford University Press.

Frege, G., 1917: "The thought." In M. Beaney, (ed.), 1997: *The Frege Reader*. Oxford: Blackwell.

Friedman, O., and Leslie, A.M., 2007: "The conceptual underpinnings of pretense: Pretending is not 'behaving as if' ." *Cognition* 105, 103-24.

García-Ramírez, E., forthcoming: "An integrated account of substitution failure".(ms). Philosophy Department, University of Michigan.

García-Ramírez, E., and Shatz, M., forthcoming: "On problems with descriptivism: Psychological assumptions and empirical evidence." *Mind & Language*.

Jacob, P., 2003: "Intentionality." In E. Zalta, (ed.): *Stanford Encyclopedia of Philosophy*, http://plato. stanford.edu/entries/intentionality/

Kripke, S., 1979: "A puzzle about belief." In A. Margalit, (ed.) 1979: *Meaning and Use*. Dordrecht: Synthese.

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

Kroon, F., 2004: "Descriptivism, pretense, and the Frege-Russell problems" *Philosophical Review* 113, 1-38.

Leslie, A.M., 1987: "Pretense and representation: the origins of "theory of mind" ." *Psychological Review*, 94, 412-26.

Leslie, A.M., 1994: "Pretending and believing: issues in the theory of ToMM." *Cognition*, 50, 211-38.

Leslie, A.M., 2002: "Pretense and representation revisited" in N.L. Stein, P.J., Bauer, M., Rabinowitz, (eds.) 2002: *Representation, Memory and Development: Essays in Honor of Jean Mandler*. Mahwah, NJ: Lawrence Erlbaum.

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

Lillard, A.S., 2002: "Just through the looking glass: children's understanding of pretense." In R.W. Mitchell, (ed.) 2002: *Pretending and imagination in animals and children*. Cambridge: Cambridge University Press

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

Marcus, Ruth B., 1981: "A proposed solution to a puzzle about belief." *Midwest Studies in Philosophy* 6, 501-10.

McKay T., and Nelson, M., 2005: "Propositional attitude reports." In E. Zalta, (ed.):*Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/entries/ prop-attitude-reports/

Nichols, S., and Stich, S. 2000: "A cognitive theory of pretense." *Cognition*, 74, 115-47.

Nichols, S., and Stich, S. 2003: *Mindreading: An integrated account of pretense, self-awareness and understanding other minds.* Oxford: Oxford University Press.

Plantinga, A., 1978: "The Boethian compromise." In A. Plantinga, 2003: *Essays in the Metaphysics of Modality*. Oxford: Oxford University Press.

Quine, W.V.O., 1980: "Reference and modality." In W.V.O., Quine, 1980: *From a Logical Point of View*, Harvard University Press.

Recanati, F., 1993: *Direct Reference*. Oxford: Blackwell

Russell, B. 1905: "On denoting." In A.P. Martinich, (ed.) 2002 *The Philosophy of Language*. Oxford: Oxford University Press.

Salmon, N., 1986 *Frege's Puzzle*. Ridgeview Publishing, USA.

Salmon, N., 1989: "Illogical belief." *Philosophical Perspectives* 3, 243-85.

Salmon, N., 2006: "The resillience of illogical belief". *Nous* 40, 2, 2006, 369-75.

Schiffer, S., 1987: "The 'Fido'-Fido theory of belief." *Philosophical Perspectives* 1, 455-80.

Schiffer, S., 2006: "A problem for direct reference theory of belief reports." *Nous* 40, 2, 361-68.

Soames, S., 1989: "Direct reference, propositional attitudes, and semantic content." *Philosophical Topics* 15, 44-87.

Soames, S., 2002: *Beyond Rigidity*. Oxford: Oxford University Press.

Stalnaker, R., 1978: "Assertion." In R. Stalnaker, 1999: *Context and Content*. Oxford: Oxford University Press.

Stalnaker, R., 1987: "Semantics for belief." In R. Stalnaker, 1999: *Context and Content*. Oxford: Oxford University Press.

Swanson, E., 2006: "Frege's puzzle and pragmatic presupposition." In E. Swanson, 2006: *Interactions with Context*. PhD Dissertation. MIT.

Wellman, H.M., and Miller, J. G., 2008: "Including deontic reasoning as fundamental to theory of mind." *Human Development* 51, 105-35.

Wettstein, H., 1989: "Turning the tables on Frege or how is it that "Hesperus is Hesperus" is trivial?" *Philosophical Perspectives 3, Philosophy of Mind and Action Theory*, 317-39.

Chapter 4

An Integrated Account of Substitution Failure

Most "referentialist" accounts of the phenomenon of substitution failure (e.g., Stalnaker 1978, Salmon 1986, and Soames 2002) have focused either on the "mind" or the "language" side of the phenomenon. A proper account, however, must include both the linguistics and psychology of substitution failure. The goal of this paper is to offer such an account. The account is based on: (i) psychological data from studies on name processing; and (ii) linguistic data concerned with belief-report sentences and predicate-meaning-transfer. The psychological account shows why and how the human mind may fail to identify coreferential names and the linguistic data explains how speakers may use belief-report sentences to target such phenomena. This integrated account is tailor-made to fit a referentialist semantics for proper names.

Introduction

The referentialist theory of names claims that names contribute only their referent to the content of the sentence where it is used. There are many good philosophical (see Kripke 1980, Donellan 1960 and 1970, Putnam 1975, and Marcus 1961), linguistic (see Stalnaker 1978, Recanati 1993, and Swanson 2006), and psychological (see García-Ramírez and Shatz forthcoming) reasons to think that this theory is correct. Yet, the theory is not without problems. A major one of these is presented by the phenomenon of substitution failure.

There is substitution failure whenever the substitution of coreferential expressions fails to preserve truth value. If the referentialist theory is correct, coreferential names make exactly the same semantic contribution; thus, they should affect truth values in exactly the same way. This is consistent with the truth of (1a)-(1c) but not with that of (2a)-(2b). Referentialism is, at first glance, inconsistent with substitution failure.

(1a) Dr. Zemph wrote *Psychoanalysis.*
(1b) Clare Quilty wrote *Psychoanalysis.*
(1c) Dr. Zemph is Clare Quilty.
(2a) Humbert believes that Dr. Zemph wrote *Psychoanalysis.*
(2b) Humbert does not believe that Clare Quilty wrote *Psychoanalysis.*

Several philosophers (see Salmon 1986 and 1989, Soames 1989 and 2002, Stalnaker 1987, Crimmins 1998, Crimmins and Perry 1989, and Recanati 1993 among many others) have tried to show how substitution failure can be the result of non-semantic mechanisms, thereby saving the day for a referential semantics. All of them, however, have focused on the linguistic features of substitution failure: explaining, for example, the behavior of subordinate clauses, attitudinal verbs, semantic pretense, and presupposition carrying expressions. As a result, they have either speculated on, or not dealt with, an important part of substitution failure: the explanation of the cognitive mechanisms that give rise to the phenomenon. In this paper I will offer some such explanation. Doing so will not only complete the picture but will also help us find a more adequate linguistic account of the phenomenon.

I start, in section 4.1, by describing the phenomena in need of explanation. I will argue that, strictly speaking, there is substitution failure only within reports of attitudinal mental states (e.g., 'John believes that . . . '), though there seem to be uses of other sentential constructions in equal need of explanation. In section 4.2, I present some psychological data on the processing of names. I focus mainly on evidence concerning the mechanisms of storage and retrieval of proper names, but will also make use of studies on the semantic assumptions that play a role in the acquisition of proper names. Based on this evidence, section 4.3 offers the psychological side of my theory by showing why and how speakers may fail to realize that two names are coreferential. Section 4.4 completes my theory by describing a linguistic mechanism that may help us better understand belief reports that exhibit substitution failure. In section 4.5 I apply the integrative part of the theory by showing how it can solve Frege's famous problem of informativeness as well as Kripke's puzzle about belief. Finally, section 4.6 offers some extra support to the theory by showing how it can be used to solve some further puzzles.

4.1 How much substitution failure is there?

I said that there is substitution failure whenever the substitution of coreferential expressions fails to preserve truth-value. The tradition has generally focused on propositional attitude ascriptions. To illustrate, consider the sentences in (1) and (2), all of which are true.

(1a) Dr. Zemph wrote *Psychoanalysis.*
(1b) Clare Quilty wrote *Psychoanalysis.*
(1c) Dr. Zemph is Clare Quilty.
(2a) Humbert believes that Dr. Zemph wrote *Psychoanalysis.*
(2b) Humbert does not believe that Clare Quilty wrote *Psychoanalysis.*

Some seem to think there is substitution failure also among simple sentences such as those in (3). The examples are owed to Saul 1997.

(3a) Clark Kent came into the phone booth, and Superman came out.
(3b) Dan dresses like Superman.
(3c) Superman was more successful with women than Clark Kent.

Others seem to think there's also substitution failure within epistemic contexts that are not mental state reports, such as (4a) which is allegedly true and (4b) which is false (both sentences are uttered by E in Ann Arbor).

(4a) It is a priori that I am here.
(4b) It is a priori that E is in Ann Arbor.

Strictly speaking, only the examples involving attitude ascriptions exhibit substitution failure. Let me explain why. First, we know, as a matter of fact, that Superman is an F iff Clark Kent is an F. Hence if we want to accept that there is substitution failure in (3a)-(3b) and that (3c) is true, it *must* be that 'Superman' and 'Clark Kent' are not being used as coreferential terms. If we insist, however, that the names are used coreferentially, then we must accept that (3a)-(3b) do not exhibit substitution failure and that (3c) is false. There seems to be no more evidence to consider in these cases. Speaker intuitions may argue against this latter move, but these intuitions seem compatible with the claim that the names involved are not being used coreferentially.[1] Either way, there seems to be no substitution failure in (3). Notice, however, that we cannot make the same claim about (2a)-(2b), for even though we know that Dr. Zemph is an F iff Clare Qulity is an F, we have independent reasons to accept (2a)-(2b) as true. In this case, there is extra evidence to consider. We know, because we have observed his behavior, that Humbert doesn't believe that Clare Quilty wrote *Psychoanalysis*. Attitude ascriptions seem more puzzling.

Second, there are good reasons to think that either the sentences uttered by E in (4a) and (4b) are not synonymous or that both (4a) and (4b) are false. It seems true that whenever a speaker utters the English sentence 'I am here' she manages to say something true. So it seems appropriate to say that competent English speakers know (in some broad sense of 'knows') that every utterance of 'I am here' expresses a true proposition. The best way to

---

[1]The comic-book story suggests that Superman and Clark Kent are two temporal parts of the same individual.

put this is as follows: it is *a priori* that an utterance of 'I am here' is true. Now, one thing is the truth-value of a proposition expressed in a context and another thing is the proposition itself. There is a big step between accepting that competent English speakers know *a priori* the truth-value of a given utterance and accepting that competent English speakers know *a priori* which proposition it is they know the truth-value of. To know that by uttering 'I am here' I'll be expressing the proposition *that E is in Ann Arbor* I need to know that E is the speaker of the utterance and that Ann Arbor is the place of the utterance, and there is no *a priori* way to know any of these. That is exactly why (4b) is false. This leaves us with two options, either E asserts that it is *a priori* that E is in Ann Arbor by uttering (4a), or E asserts that it is *a priori* that his utterance of 'I am here' is true. If the former is the case, then (4a) is false and so it has the same truth-value as (4b). If the latter is the case, then the use of 'I' and 'here' in (4a) are not coreferential with the use of 'E' and 'Ann Arbor' in (4b). Either way, there's no substitution failure here.

I agree with Saul 1997, nonetheless, that there is something puzzling about (3), specially if we also accept that such sentences may be uttered by speakers who know that Superman is Clark Kent. Since I think a proper account of the sentences in (3) takes 'Superman' and 'Clark Kent' to be used as having different referents, I still need a story that explains how a competent speaker may use such terms in such ways in light of her knowledge, and how she may successfully communicate this use to her audience. The account of substitution failure I want to present should be good enough to illuminate this.

4.2   The evidence

How is it that natural language speakers process names? Which limitations are in place when speakers *acquire* new proper names? To answer these questions I will look at studies on memory and cognitive development. The evidence suggests that certain accounts of substitution failure are misguided and, as I will argue, it also suggests an alternative resolution of Frege's puzzle.

4.2.1   Memory

Research on memory has suggested that proper names are uniquely and arbitrarily linked with faces in semantic memory (see Semenza and Zettin 1989, Semenza and Sgaramella 1993, Valentine, Brennen, and Brédart 1996, and Werheid and Clare 2007; for related research on names of places see Goodglass and Wingfield 1993). The processing models

associated with these studies suggest that no descriptive information about the referent of the name is necessary for proper name processing.

Names are difficult to remember. Everyone has, at some point, been unable to recall someone's name, while knowing pretty well who that person is. This problem is, as Kaplan puts it, "a looming fact of life" [1990, p.105]. The speaker is acquainted with the referent, she might be able to describe the referent, she also knows the name, and yet the name is not retrieved. Folk-psychology describes this as the tip-of-the-tongue or TOT phenomenon.

Cohen 1990 tested the differences in processing meaningful and meaningless expressions. The results show that recalling proper names is just as difficult as recalling meaningless, non-word expressions (e.g., 'wesp', 'blick'). Based on this evidence, Cohen argues that the TOT problems with proper names are ascribable to a lack of semantic associations. On her view, proper names are detached from conceptual representations. There are different interpretations of the evidence, of course, but there is general agreement that proper names are vulnerable in terms of retrieval and that this is owed to the fact that "a proper name conveys almost no information about the entity it names" [Valentine et.al. 1996, p. 108]. This is not to say that they provide no information whatsoever. A proper name like 'John', for example, may convey the information that the referent is English, or of English origin. The important claim is that, according to Valentine et.al. 1996, proper names do not convey any *uniquely identifying* information that could allow us to single out their referents.

This point very clearly suggests that proper names are not, in any way, *reliably* associated with files containing descriptive information about their referents. This goes against all theories that intend to explain substitution failure in terms of descriptive information that is supposed to be both conveyed by the use of proper names *and* truth-conditionally relevant in belief reports. This is the case even for some "referentialist" accounts such as Recanati's 1993. On his view, names contribute their modes of presentation (MOFPS) to the truth conditions of belief-reports, each name having a different MOFP. These MOFPS are supposed to be mental files in the form of encyclopedia entries that include descriptive information about the referent of the name. The theory (see Recanati 1993, p.184-187) presupposes an intimate relation between the linguistic and the psychological modes of presentation. That is how names are supposed to contribute the descriptive information associated with their psychological MOFP in a way that is truth-conditionally relevant within belief reports.[2]

---

[2]On Recanati's own words (or almost): "Suppose someone says (2a). If the hearer contextually assumes that the ascriber is being 'faithful' to the believer, he will be led to assume that the ascribed belief concerning Dr. Zemph involves a mode of presentation [...] which includes the descriptive concept 'called "Dr. Zemph".' Hence, the hearer will assume that the believer thinks of Dr. Zemph as 'Dr. Zemph'; he will take the reference of the 'that'- clause as a quasi-singular proposition involving a certain (*de re*) concept over and beyond the reference of [...] 'Dr. Zemph' " [Recanati 1993, p.358].

Theories like this are undermined by the empirical evidence and appear to be inconsistent with what cognitive psychologists have to say about the latter.

In most models (see Valentine et.al., 1996 for a detailed survey) the files containing descriptive information about the referent do not play a necessary role in name retrieval. Speakers can access such descriptive information about the referent of the name, yet this is *not* necessary for comprehension and production. If such were the case, as Recanati's model predicts, then there would be more entry points, within memory, from which to access a name. If so, then proper names should not be specially difficult to retrieve.[3] The evidence clearly shows that this is not the case. Descriptive information, of the sort we find in an encyclopedia entry, is stored in separate and independent storage spaces from that of the name. Studies on anomia seem to confirm this while suggesting that names are dissociated from descriptive information. For a detailed account of the neuropsychology of proper names see Semenza 2009.

Furthermore, studies on lexical and cognitive development suggest that infants as young as 6 months of age show comprehension of their own names while still being unable to process descriptive information such as *being called 'N'*. For a more detailed discussion of these issues and the relevance of these data for the philosophical debate on proper names see García-Ramírez and Shatz forthcoming.

The central point is clear: in order to access a proper name within memory it is *not* necessary to *also* access descriptive information about its referent. That is why names are difficult to retrieve and why there is selective impairment and preservation of names (i.e., nomic aphasia, see Semenza 2009). There is, briefly put, no intimate relation between the proper name and the descriptive information that it may be associated with. In some cases (e.g., early infancy and anomia) there might not even be any descriptive information associated with the name. Without this intimate relation, accounts of substitution failure that appeal to descriptive information that is somehow contributed by the name seem implausible. They fail on psychological grounds.

4.2.2 Semantic assumptions for proper names

Hall 2002 reviews different studies on how semantic knowledge aides proper name acquisition. He considers four different assumptions, two of which are of great importance for our discussion: first, speakers presuppose that there is only one object per name; and, second,

---

[3]Note that the claim that the descriptive information may be "arbitrarily" associated with the name makes no difference here. If the information is to be always associated with the name, as Recanati claims, it must always offer extra entry point from which to access the name. This would make names easier to retrieve.

speakers presuppose there is only one name per object.[4]

One object per name

Hall 1996 conducted a study designed to show that children as young as 4 years of age have the default assumption that in order for something to be a proper name it must apply to only one object. Hall divided the subjects (ninety 4-year-olds) in two groups. Both groups heard a novel word, e.g., "ZAVY", modeled in an ambiguous sentence context, e.g., "This dog is …", that supported both *proper name* and *adjective* interpretations of the novel word. Group one heard the novel word applied to a drawing of a single object with a salient property (e.g., a striped dog). Group two heard the novel word applied to a drawing of two objects with the same salient property (e.g., two striped dogs). The study predicted that children in group one would interpret the novel word as as a name, while children in group two would take it to be an adjective.

To test this prediction children in both groups were presented with three sets of drawings including (a) the labeled object or objects, (b) an object of the same kind as the labeled one but lacking the salient property (e.g., a non-striped dog), and (c) an object of a different kind with the same salient property (e.g., a striped umbrella). The study then assessed how the children in both groups would extend their interpretations of the novel word by asking questions that used the novel word in ambiguous ways (i.e., consistent with both proper name and adjective interpretations). The experimenters would ask: "Is this dog ZAVY?" or "Is this umbrella ZAVY?" The results confirmed the hypothesis that 4-year-olds assumed that proper names can only apply to single objects. Children in group one refused to extend the use of the novel word to apply to more than one object, while children in group two did not.

In a different study, Hall 1996 also showed that the "one object per name" assumption is a *default* one which can be *revised* given enough evidence. In other words, children initially refuse to accept that two objects (e.g., two different kids) may have the same name, but this assumption can be revised given enough evidence. To show this Hall presented the subjects with unambiguous proper name use of novel words, e.g., "This dog is named ZAVY", and assessed their interpretations of the words. Children in both groups made a proper name

---

[4]A third assumption, "salient individuals", was first noted by Katz, Baker, and Macnamara 1984 in a study with 3 year olds. They noted that children's acquisition of novel words varied according to the sentence environment in which they were presented, when applied to human-like objects (e.g., a doll). Proper name interpretations followed syntactic contexts like 'This is X' whereas adjective interpretations followed contexts like 'This is an X' (see also Bloom, 1990). This restriction, however, did not apply when the novel word was used to refer to objects that were not human-like (e.g., a colored block). A further study by Gelman and Taylor 1984 replicated this finding. This suggests that speakers would normally interpret a novel word as a proper name if it applies to *salient* individuals.

interpretation even though they heard the novel word applied to more than one object. The evidence suggests that, unless given evidence to the contrary, speakers presuppose that to each name corresponds only one object. Anecdotal evidence (see Macnamara 1982), showing that children are in fact reluctant to accept that two objects (e.g., two friends) can have the same name, supports this claim.

Hall's 1996 study shows evidence on behalf of two different claims. First, that when interpreting "ZAVY" in an ambiguous environment "This dog is...", children will interpret the word as a proper name if the context singles out a salient referent. Children will interpret the novel word as an adjective if the context does not single out a salient individual. This suggests that when interpreting a novel word in a syntactically ambiguous environment, speakers will take it to be a proper name if the context also provides enough evidence to single out the word's referent. In the case of the experiments this is achieved by presenting the subjects either with a drawing of a single dog or one of multiple dogs. Second, the study also shows that, once they have interpreted the novel word as a name, children will refuse to extend its use to refer to other individuals. This assumption may be revised, as Hall shows, when given enough evidence to the contrary.

Another way to put these claims is to say that in contexts that provide a single salient referent, speakers normally interpret novel words as proper names *and* that whenever interpreting a word as a name speakers will not extend it to refer to multiple objects. This naturally prompts the question: why do speakers go from the evidence of a single referent to a proper name interpretation? Hall's suggestion is that speakers normally assume that there is only one object per name.[5]

One name per object

If speakers presuppose that to each name corresponds only one object, it seems natural that they also presuppose that, if they encounter a new name, it must belong to a different, previously unnamed, object. Hall and Graham 1999 conducted a different study intended to show that children do aid themselves with this assumption; i.e., they assume that to each object corresponds only one name.

Forty-eight 4 and 5-year-olds and forty eight 3-year-olds heard a novel word as it was applied to a familiar stuffed animal (e.g., a stuffed dog). Subjects were divided in two

---

[5]In conversation, Eric Swanson has mentioned that these claims would be in better standing if it could be shown that children prefer the proper name interpretation *even* in plural environments such as "These dogs are ZAVY and LAVY". It seems to me that such experiment could, in principle, offer some such support. Yet I have some reservations about its implementation. In any case, I leave the job of designing such experiments to psycholinguists. Suffice it to say that, as far as I know, the evidence available in the literature does not go against Hall's claims.

groups. Group one heard the novel word (e.g., 'X') modeled syntactically as a proper name (e.g., "This dog is named 'X' "). Group two heard the word modeled as an adjective (e.g., "This dog is very X"). All subjects were then presented with a second stuffed dog. They were asked to determine which of the animals was the referent of a second novel word (e.g., 'Y'). Both groups were then divided in half: one half of each group heard the second word modeled as a proper name (e.g., "Show me a dog that is named 'Y' "); the remaining half heard the second word modeled as an adjective, e.g., "Show me a dog that is very 'Y' ". Once the subjects decided which object was the referent of the second new word, they were asked whether the word in question could also apply to the object they did not choose.

Hall and Graham predicted that the choice of reference for the second novel word would vary across groups. In particular, they predicted that children would reject having two names for the same object. Thus, subjects in group one who heard the second novel word 'Y' modeled as a proper name would take it to apply to the second stuffed dog and reject the possibility that it apply to the other, previously named, dog. A contrasting claim was made about those (group two) who heard the first novel word modeled as an adjective. Hall and Graham predicted that children would *not* reject applying two adjectives for the same object. In particular, they predicted that children would not be driven to apply a second adjective to the new, unnamed, object. And, when they did apply another adjective to an unnamed object, the subjects would *not* refuse to extend the second adjective to the previously named object.

The results confirmed their general prediction that children's interpretations were "significantly affected by the lexical form classes of the two novel words" [Hall and Graham 1999, p.88]. This use of lexical categories was particularly strong with respect to proper names, suggesting that children presupposed that new names only apply to previously unnamed objects or, alternatively, that to each object corresponds only one name.

> [I]f both words were modeled as proper names, children showed the strongest tendency to select the unlabeled object as the referent of the second word and, having chosen that object, to deny that the second word could apply to the already-labeled object (e.g., as if one dog could not be both "Fido" and "Rover") [Hall and Graham 1999, p.88].

The studies presented in Hall 2002 suggest that, from a very early stage of development, speaker knowledge of proper names includes two default assumptions that can be dubbed as "unique reference" and "unique naming". According to the former, a single proper name refers to only one object—I deal with homonyms in 4.3.2. According to the latter, individual objects have only one name or, better put, different names refer to different objects. To

say that both assumptions are "default" is to underscore the fact that, when given enough evidence to the contrary, speakers do revise their assumptions and accept the exceptions.

Here's some anecdotal evidence that supports the claim that these assumptions carry on through development until adulthood. The *National Gallery* in London dedicates one hall to French impressionism. It exhibits, among other things, some four or five paintings by Monet. All but one of these paintings are attributed to some Claude O. Monet. The remaining one is attributed to a Claude Monet. My wife, who is an admirer of Monet's work, was surprised. She confessed to wonder whether there were two Monets, Claude O. and Claude. Her doubt was, I think, completely justified. Why else would a curator use two names, 'Claude' and 'Claude O.' for the same artist?

What is important for the philosophical discussion is what these assumptions tell us about what speakers normally do when acquiring two different names that, by accident or not, have a common referent: i.e., the scenario where substitution failure appears. The empirical evidence suggests that speakers start by assuming that a given name, say *m*, refers to only one individual and that such name *m* will be enough to refer to that individual in all circumstances. Hence, they assume as well that a different name, say *m'*, will refer to a different individual. It follows that cases where two different names, say *m* and *m'*, refer to the same object will be considered exceptions and, hence, will not be accepted unless given enough evidence to the contrary. It should not be a surprise, then, that speakers do not give the same credence to property ascriptions that differ merely in the use of different coreferential names. Thus, it should not be a surprise that these cases give place to substitution failure—it would be surprising if they didn't. What is most important is the fact that the existence of the phenomenon (i.e., of substitution failure) seems to have more to do with speaker's assumptions about names than with the *semantic content* of names strictly speaking.

As I will show in what follows, this evidence suggests an explanation of what goes on when speakers competently understand two *names* and yet fail to realize they are coreferential. Even though illuminating this psychological state of the speaker is relevant to understand substitution failure, it is not all there is to the phenomenon. There may be attitude ascriptions exhibiting substitution failure where the ascriber doesn't use names but, say, demonstratives. Furthermore there may be adequate attitude ascriptions with substitution failure where we cannot appropriately claim that the ascribee is processing or understanding exactly the same name that the ascriber is using. I am aware of this. Still, the evidence does illuminate, at least, part of the phenomenon. It is this relevant evidence that suggests an account that, as I will argue, offers the best explanation of the data.

4.2.3 The localization of labels

What happens when a speaker takes two labels to denote different objects? Anderson and Hastie 1974 tested undergraduates on processing information depending on whether it was introduced by a proper name or a definite description. Subjects were given a number of predicates ('F', 'G' and 'H') about a person A and introduced her with a proper name *'N'* such that $N = $ A. They were then given different predicates ('R', 'Q' and 'S') about the same person, but this time the person was introduced by means of a definite description *'DD'*, without the subjects knowing so, such that $DD = $ B. This generated the assumption that the proper name and the definite description denoted different objects. The subjects were then told that A = B. This forced the subjects to make inferences when asked to verify a sentence that used either term (i.e., *'N'* or *'DD'*) followed by the predicates introduced by means of the other term (e.g., *'N* IS R', *'N* IS Q', and *'N* IS S'). The subjects were also asked to verify sentences using terms and predicates that were initially introduced together. These were non-inferences (e.g., *'N* IS F') used for comparison.

Inferences with the definite description took more processing time than those with the proper name. In fact, name inferences (i.e., verifying sentences such as *'N* IS R') took the same time as name non-inferences. Description inferences (i.e., verifying sentences such as *'DD* IS F' and *'DD* IS G') took up to twice as long as description non-inferences (e.g., *'DD* IS R'). Anderson and Hastie claim that this is due, first, to the fact that when the information is initially stored as belonging to two different individuals, two memory nodes, A and B, are created. Second, they claim that when the identity A = B is established, the proper name *'N'* is directly connected with the second node B; whereas the definite description *'DD'* is only indirectly connected with node A by means of a third node A = B.

This difference in processing suggests that, when taken to be about different objects, the information is stored in unrelated storage spaces. This evidence, together with the assumptions described above (i.e., unique reference and unique naming), suggests that when presented with different but coreferential names—and without enough counter evidence— speakers will normally store the information associated with each name in different unrelated files.[6]

---

[6]From now on I will use 'file' and 'storage space' interchangeably. There is no commitment to any particular account of concepts, or memory, in my use of them.

4.3   The psychological side

The studies on memory (section 4.2.1) suggest that proper names are, memory-wise, *task specific* and *localized*. They suggest that a unique cognitive architecture and memory resources are required for proper name processing. The studies on nomic aphasia in particular (see Semenza 2009) suggest that proper names are memory-wise isolated from the rest of the lexicon. Part of this specificity of names results in their being localized: the descriptive information that may be associated with a name is stored separately. The result is an important independence of storage and retrieval. Storing and retrieving a name is *not* the same as storing retrieving descriptive information.

Given this cognitive limitation, the studies on the acquisition of names (section 4.2.2) suggest that whenever a speaker acquires two novel names that she takes to belong to two different individuals she will store these names in two unrelated files, thereby granting independent patterns of inference. Whether the names in fact are coreferential will be immaterial to the way the subject stores them in memory.

4.3.1   The account

This, to my mind, explains how a competent and rational speaker can successfully use two different names (e.g., 'Dr. Zemph' and 'Clare Quilty') and still fail to realize that they are coreferential. So long as the speaker presupposes that such names label different objects, there is nothing puzzling about it. If so, then we also have a resolution of Frege's 1892 puzzle. Many philosophers (see Recanati 1993, Crimmins and Perry 1989, Jacob 1990, Fodor 1990, Richard 1990, and Kaplan 1989) agree that any solution must meet Frege's constraint:

> [I]f *x* believes *y* to be F and also believes *y* not to be F, then there are distinct modes of presentation *m* and *m'* such that *x* believes *y* to be F under *m* and disbelieves *y* to be F under *m'*. [Schiffer, 1978, p. 180]

The empirical evidence I have presented proves its philosophical worth by suggesting an alternative interpretation of Frege's constraint, what I dub "storage constraint". On this reading, there is no need of descriptive information that must be associated with names in any way.

storage constraint:

> [I]f *x* believes *y* to be F and also believes *y* not to be F, then there are distinct **names** *m* and *m'* stored in different **files**, such that *x* believes *y* to be F when associated with *m* and disbelieves *y* to be F when associated with *m'*.

This account makes use of two central notions: DIFFERENT NAMES and DIFFERENT FILES. The latter seems less confusing. Two given sets of information are stored in different files whenever access to them generates similar patterns of inference as those studied by Anderson and Hastie 1974 as described above. The important contribution of this notion, i.e., DIFFERENT FILES, is that the information so stored is inference-wise unrelated, which explains why the subject can ascribe incompatible properties to objects associated with different files. The other notion, i.e., DIFFERENT NAMES, is less clear and seems to get us into trouble.

There may be, as I said, attitudinal ascriptions with substitution failure involving demonstratives and not proper names (e.g., "Peter thinks that we are *here* not *there*"). It may be objected that the evidence I have given on proper names is irrelevant for these cases since it is about proper names and not demonstratives. Let me now clarify my argument. I do not take my account to be the only one consistent with the evidence. Rather, I take it to offer the best explanation of the available evidence, which includes psychological data on proper name processing as well as linguistic data on substitution failure. The lack of equivalent psychological data on demonstratives is not itself a problem. In so far as there is no evidence against it, the account stands on the available evidence. (This theory will also, of course, rest on its explanatory power; for more on this see section 4.6.) I believe the resulting theory offers a satisfactory explanation of the mental states that underlie substitution failure. The outcome is something like **storage constraint***:

storage constraint*

> [I]f *x* believes *y* to be F and also believes *y* not to be F, then there are distinct **referential expressions** *m* and *m'* stored in different **files**, such that *x* believes *y* to be F when associated with *m* and disbelieves *y* to be F when associated with *m'*.

This formulation, however, is still problematic. There may be adequate attitude ascriptions that exhibit substitution failure in cases where the ascribee may not properly be said to understand exactly the same referential expressions used by the ascriber. Suppose you're trying to describe Pierre's behavior and that Pierre is a monolingual French speaker. You might as well say something like "Pierre believes we are *here* (pointing to A on a map) not *there* (pointing to A on a map)". It seems inappropriate to say that Pierre stores 'here' and 'there' in unrelated files. Or consider a more hackneyed example. Hammurabi, a famous ancient astronomer, believed that Hesperus but not Phosphorus was visible in the evening. Yet, we can be pretty sure that Hammurabi didn't understand either 'Hesperus' or 'Phosphorus'.

To avoid getting into trouble with these cases we must make certain assumptions explicit. Notice, first, that all attitudinal ascriptions (in general) must meet the following simplified correspondence principle if they are to be acceptable (let alone true):

correspondence:

> *S* will successfully ascribe the attitudinal state of Fing that R to *Y* by using a sentence of *L* of the form "Y F'S that R" where 'Y' is a referential term, 'F' is an attitudinal verb and 'R' is a sentence of *L* of the form "N is P" where 'N' is a referential term and 'P' a predicate, iff *S's* use of 'N' somehow corresponds to the mental representation that the ascribee associates with the property denoted by 'P'.[7]

Since it is an assumption of the debate that the attitudinal ascriptions with substitution failure are not only acceptable but also true, **correspondence** makes explicit an important assumption of the debate. If we take this assumption into account, we get the **official** formulation of the **storage constraint**.

storage constraintOFFICIAL

> [I]f *x* believes *y* to be F and also believes *y* not to be F, then there are distinct **corresponding referential expressions** *m* and *m'* stored in different **files**, such that *x* believes *y* to be F when associated with *m* and disbelieves *y* to be F when associated with *m'*.

I believe that **storage constraint**OFFICIAL delivers an adequate psychological account of the relevant mental state of the ascribee in an attitudinal ascription with substitution failure. But before we can have the whole account, we need to clarify what counts as distinct referential expressions. To simplify things, I will focus on proper names, but the criteria I will offer should be applicable to referential expressions in general.

---

[7]I owe this suggestion to Lenny Clapp. The principle is simplified because it is silent about cases where attitudinal ascriptions are expressed by using complex sentences in the embedded 'that...' clause. The formulation should be clear enough to see how it can be extended to cover these latter cases. If the reader worries about the phrase "somehow corresponds" here's one way in which it can be understood. To say that *S's* use of 'N' in her ascription of *Y's* attitudinal states corresponds to the relevant mental representation in *Y's* mind is to say that 'N' is the referential term that *Y* would use to report her own state if she were a competent user of *L*. Or alternatively, to say that there is such correspondence is to say that if *Y* were a competent monolingual speaker of *L* the relevant representations she would associate with the property denoted by 'P' would correspond to the object she would refer to by 'N'.

4.3.2 How to count names

How do speakers go on individuating names? I think the answer to this question is fairly intuitive, though subtle. I will present three different criteria that may generally guide speakers in individuating names. These criteria need not be infallible. The view I'm about to defend is a close cousin of Kaplan's 1990 "common currency" view of names. Yet, as you will see later on, our views differ importantly when it comes to explaining substitution failure.[8] There is an intuitive way to distinguish between names. This is the way in which 'Dr. Zemph', 'Clare Quilty', 'John' and 'George' are all distinct, and the way in which 'John' and 'John' are not. I will call this the "phonographer's" criterion.

phonographer: two strings of symbols or sounds count as two different referential expressions if the strings differ phonemically.

Although very intuitive, this criterion alone is insufficient, since it fails to capture a distinction that seems equally important: that between homonymous names. Speakers do presuppose (see 4.2.2) that names have only one referent. But, as a matter of fact, different referents may share the same string of sounds as their names: e.g., George Bush and George Bush, Aristotle Onasis and Aristotle the Stagirite. This would not be a problem if we did not have good reasons to distinguish between homonymous names. But we do have such reasons. Speakers may not only, and intelligently, *utter* sentences that require some such distinction, e.g.,'George Bush is not George Bush' or 'Aristotle did not write Aristotle's works', they may also have mental states that seem to make the same requirement. In so far as speakers may associate different (and sometimes incompatible) sets of information under what phonemically counts as the same name, we need to count homonymous names as different names.

There is, then, a second intuitive sense in which 'George Bush' and 'George Bush' are two different names. I will call this the "reference" criterion.

reference: two given strings of symbols count as different referential expressions if they have different referents.

Clearly, this criterion alone is not enough. By itself, it does not help us distinguish between coreferential names like 'Dr. Zemph' and 'Clare Quilty' or 'Tully' and 'Cicero'. These distinctions are not only intuitive but necessary if we are to make sense of the storage constraint. Speakers naturally take names that differ phonemically or orthographically to be

---

[8]First, because Kaplan 1990 merely hints at some such explanation; he does not develop it. Second, because Kaplan thinks we must "attack a version of Frege's principle of compositionality" [1990, p.95]. The theory I here present does not follow any such strategy.

distinct and it may be in virtue of this that they end up storing the information associated with each name under different files. It seems, then, that we need both criteria. Thus, the reasonable thing to do is to take both criteria as offering, each of them, conditions for two strings of symbols to count as different names.

Finally one last criterion owed to Kripke 1980 and Kaplan 1990: the "history" criterion.

history: two phonemically identical and coreferential expressions count as different if they result from two distinct historical events that coincide, phonemically and referentially, by sheer accident.[9]

Now, if we take these criteria to offer generally trustworthy conditions for two strings of sounds (or symbols) to count as two different names, the result is what I call "difference":

difference: two strings of symbols count as different names either:

δ_1: because they differ phonemically; or
δ_2: because they differ referentially; or
δ_3: because they differ historically.

According to these criteria, speakers will normally take any two strings of symbols that meet either one of δ_1, δ_2, or δ_3 to count as different referential expressions. This is good enough for the purposes of the **storage constraint**$_\text{OFFICIAL}$.

I am aware that these criteria *may* have odd metaphysical consequences, though I would be skeptical about drawing such conclusions.[10] What matters here is how speakers count names. The metaphysical nature of this individuation may be uninteresting. Human understanding is what is of interest here.[11]

Briefly put

On this view, a rational and competent speaker may fail to notice that two distinct expressions are coreferential, even though she may know what both refer to, because she has

---

[9] On Kripke's 1980 words: "two totally distinct 'historical chains' that by sheer accident assign phonetically the same name to the same man should probably count as creating distinct names despite the identity of referents. The identity may well be unknown to the speaker, or express a recent discovery." p. 8 ft. 9.

[10] It is a consequence of these criteria that if two instances of the phonetic string 'David' are used to refer to different objects they count as different names. Some have objected to this conclusion (see Kaplan, 1990). On their view, David Hume and David Kaplan "have the same name". I think it is useful to distinguish between names and naming traditions and to notice that it is only the former, not the latter, that is semantically relevant. David Hume and David Kaplan have the same name as David Lewis in the sense that they are all part of the same old western tradition of using the string 'David' to name male individuals. But it's not true that David Hume, David Kaplan, and David Lewis have the same name in the sense that we can use this common name as a singular term to refer to any of them, or to successfully substitute David Hume's name for David Kaplan's. See Kaplan 1990 for a similar yet alternative distinction in terms of "generic" names.

[11] For a similar approach toward drawing metaphysical conclusions from cognitive accounts, see Leslie 2008, p. 34.

stored them in different files. This is the psychological account of substitution failure that I favor. I do so, at least, for three reasons: first, it is empirically substantiated (in the sense described); second, it is very cheap cognitively (i..e, it does not make any extraordinary cognitive demands on the side of speakers); and third, it helps us solve all the relevant puzzles.

4.4   The linguistic side

Several philosophers (see Salmon 1986 and Crimmins and Perry 1989) have underscored the similarity between 'that...' clauses and quotations. Just like the latter, the former may refer either to contents of the words within them or to the words themselves. Recanati 1993 argues, convincingly I think, that there is one more respect of similarity. Like quotations, the reference of 'that...' clauses is underdetermined in much the same way as that of a demonstrative. There is no linguistic rule determining the reference of my use of 'he'. To find out what the speaker refers to by her use of 'he' the hearer must ask what the speaker can plausible be referring to. Something similar happens with quotations and 'that...' clauses. Quotations may or may not refer to the sentence-type within them; 'that...' clauses may or may not refer to the proposition expressed by the sentence within them.

According to Recanati 1993 any resolution of this underdetermination must meet two constraints:

top-down: the interpretation must be a plausible candidate for the status of content of the ascribed belief or speech act.

bottom-up: the interpretation must be recoverable from the embedded sentence.

The important point, claims Recanati, is that the interpretation process involved in the bottom-up constraint may involve what he calls "optional p-processes" which are "like enrichment or transfer which take the interpreter from the basic semantic value of a sub-expression to some further non-basic value" [p. 356]. Recanati describes one such process that makes use of descriptive information supposed to be intimately related with proper names. As I said before, this proposal lacks empirical support. But there is, at least, one other enrichment or transfer process that can help us out here: predicate meaning-transfers.

4.4.1 Meaning transfers

Speakers commonly use expressions to convey something different from its ordinary or conventional content. Names of people ordinarily refer to their bearers, but they seem to be also used to refer to other objects that stand in a salient relation to them. Names of writers, for example, are commonly used to talk about their work. So we can say, in a seminar on German Idealism, that "Kant is very obscure and confusing". In so doing, we do not intend to refer to Emmanuel Kant himself but to his work, and the common ground of the conversation includes a salient relation between the german philosopher and his philosophical oeuvre. Similar cases of deferred interpretation are illustrated in the examples (owed to Nunberg 1979) in (5).

(5a) Dr. Zemph enjoyed hearing himself read aloud.
(5b) I am parked out back.
(5c) The chair you're sitting in was common in nineteenth century parlors.
(5d) Caedmon, who was the first Anglo-Saxon poet, fills only a couple of pages of this book of poetry.
(5e) The ham sandwich is waiting for his check.

Nunberg 1993 and 2004 distinguishes between deferred *interpretation*, a phenomenon mainly concerned with predicates, and deferred *reference*, a phenomenon mainly concerned with indexicals and other referential expressions. Based on evidence on sortal crossings, Nunberg argues that the examples in (5) are best understood as cases where the meaning transfer is applied to the predicate. In (5b), for example, the predicate 'being parked out back' gets interpreted as "being the owner of a car that happens to be parked out back". This lets us maintain coreferentiality in constructions that ordinarily impose conditions of identity, like those in (5a) and (5d).

I follow Nunberg 1979 and 2004 in taking these phenomena to be the result of a mechanism of "meaning transfer" by means of which the speaker lets the hearer identify a property ψ as the correct interpretation of a predicate 'P' that ordinarily refers to a different property ϕ. Nunberg argues, convincingly to my mind, that whether this mechanism is semantic (i.e., lexicalized) or pragmatic will vary from case to case.[12] In any case, this distinction will be orthogonal for our purposes. The proposal I am going to present depends on there being meaning transfer *on the predicates*, not the names, involved in cases of substitution failure.

---

[12]"[T]his is not a simple matter of privative opposition: the mere fact that an alternation is in some way conventionalized does not mean that it no longer has any pragmatic basis. In the extreme case, alternations may be preserved in the lexicon as the *disjecta membra* of transfers that no longer have any productive role in the language ... At the other end of the scale, there are uses like in the *ham sandwich* cases that are obviously extralexical—that is, as generated exclusively by pragmatic principles, with no need of any lexical specification. In the middle, however, lies a very broad range of productive alternations whose state is less clear" [Nunberg 2004, p.351-52].

Hence, we can keep a simple and homogeneous "referentialist" semantics for names.

What are the conditions for "deferred interpretation" or predicate-meaning-transfer? Nunberg 1979 and 2004 defends two different criteria: functionality and noteworthiness.

For a predicate 'P' to transfer its meaning from ϕ to ψ there must be a relation, or function, mapping ϕ to ψ. This can be in either of two ways: either because there is a relation between the properties, for example, when ϕ resembles or evokes ψ as with WOLF and RAPACIOUS; or because there is a salient relation between the bearers of ϕ and ψ, as between authors and their work. This relation must be part of the common ground of the conversation, and ψ—the alternative interpretation—must be among the "set of things that the speaker might rationally be construed as intending to refer to in a given context" [Nunberg 1979, p.157]. This gives us the "criterion of functionality":

functionality: Let Φ and Ψ be sets of properties that are related by a salient function $g_t:\phi\rightarrow\psi$. Then if P is a predicate that denotes a property $\phi\in\Phi$, there is also a predicate P', spelled like P, that denotes the property ψ, where $\psi=g_t(\phi)$. [Nunberg 2004, p.348]

What matters in these cases is what the hearer can do given: (i) the information available in the conversational context; and (ii) the information given by the speaker. For every context there are, of course, indefinitely many different functions mapping the ordinary referent to different values. The hearer must be able to pick out the correct function on the basis of the information from (i) and (ii). Because of this, Nunberg points out, the relevant function must be "contextually salient". To illustrate, consider the following example. Two friends decide to buy several bottles of wine, from the same region and same grape, to find out which year produced the best harvest. A salient function in this context is the function "x was produced in y" that maps wine bottles into years. This allows one of the friends to point at a bottle and say "I lived in Toledo then".

Now, Nunberg points out that not all functions between sets of properties, not even all salient functions, will be acceptable for predicate-meaning-transfers. In the German Idealism seminar, for example, it is acceptable to say "Kant is obscure and confusing" but not "Kant is printed on alkaline-free paper". There seems to be some pragmatic criterion limiting the kinds of properties that can be transferred. According to Nunberg, this is given by the "criterion of noteworthiness":

noteworthiness: Predicate transfer is only possible when the property contributed by the new predicate is "noteworthy", which is to say one that is useful for classifying or identifying its bearer relative to the conversational interests. [Nunberg 2004, p.349]

So how exactly is this supposed to work? Nunberg's proposal is rather simple. All we need is a function that substitutes predicates. What speakers do by deferring the interpretation of the predicates is simply plugging in a different predicate into the same spot while leaving unchanged the sentence's logical form. Substituting predicates for saliently related predicates that refer to noteworthy properties according to the conversational interests does not result in any harsh or heavy alteration of the sentence's composition or computation.

To illustrate consider (5b).

 (5b) I am parked out back.

The predicate 'BEING PARKED OUT BACK' refers to a property ϕ that cars have depending on their geographical location. It is not a property that speakers can have. The indexical 'I', however, refers to the speaker who utters (5b). According to the deferred interpretation account, what the speaker means to say by uttering (5b) is that she has the property of having her car parked out back. For this to be the case, the predicate 'BEING PARKED OUT BACK' must transfer its meaning from ϕ to ψ, a property that speakers have in virtue of owning cars with certain geographical locations. By doing so, the LF of (5b) is left unchanged. The logical spot that was left for a property like ϕ is now occupied by another referent of the same kind, a property like ψ. All this is possible because the relation between ϕ and ψ is mediated by a relation, salient in the conversational background, between speakers and the cars they own.

This gives us a good idea of how predicate transfer works (see Nunberg 1979, 1995, and 2004 for details; see Ward 2004 for an alternative account). The lesson is that there is something like deferred interpretation whenever: (i) the speaker puts the hearer in a situation where the latter can identify the relevant function mapping the ordinary interpretation on to the intended one; (ii) the intended interpretation is among the reasonable ones that the speaker may intend given a context; and (iii) the function is salient in that context.

4.4.2   Deferred interpretation and belief reports

This is my hunch: a phenomenon like deferred interpretation takes place in contexts of substitution failure, and this explains the differences in truth-value that result from substituting names. To make my point I will contrast two kinds of contexts, with and without substitution failure, associated with belief reports. It is well known that belief reports have, at least, two readings, *de re* and *de dicto*, one of which (*de re*) is *not* subject to substitution failure. First, consider a context with no substitution failure such as **context a**.

context a:

Suppose we are talking about the history of psychoanalysis, about which Professor Humbert has written a lot. He is not present in our conversation, but his views about the topic are being discussed. We are wondering, in particular, whether Clare Quilty plays an important role in the history of the discipline. We are not sure about the merits of his *Psychoanalysis* and *More Psychoanalysis*. You've read Humbert's reviews and know what he thinks of them. So you go on and say: "Humbert thinks Clare Quilty is a pivotal element in the history of psychoanalysis".

What matters for this conversation is what Professor Humbert, a specialist, thinks *of* Clare Quilty himself. The philosopher's rendition of this belief report would be something like: "Humbert thinks *of* Clare Quilty that he is [...]." And there is no difference here between that and the following alternative paraphrase: "Humbert thinks *of* Dr. Zemph that he is [...]." This suggests that in certain contexts of belief ascription the ascribee's presuppositions about the referential expressions that correspond to those used in the report—i.e., 'Clare Quilty' and 'Dr. Zemph'—are either not so relevant or not relevant at all.

Next consider a context exhibiting substitution failure, such as **context b**.

context b:

Suppose that, as we are having our discussion about the history of psychoanalysis, Professor Humbert enters the room. You and I know that Clare Quilty will be presenting his new book at the bookstore downstairs, but Humbert doesn't. After a while, you announce: "Did you know that Clare Quilty will be visiting us later on today?" Professor Humbert is surprised: "I thought it would be Dr. Zemph, presenting his new book!" You, of course, don't know what to say. Professor Humbert is such an expert that you are afraid of correcting him. In order to ease your bewilderment I say to you: "Humbert does not believe that Clare Quilty wrote *Psychoanalysis*".

It seems clear that my utterance *cannot* have a *de re* reading. It is common knowledge to both of us that Clare Quilty wrote *Psychoanalysis*, and that Professor Humbert knows who is the author of the work. There is, however, an important difference between **context a** and **context b**. In **context a** the question under discussion is the history of psychoanalysis and not Humbert's thoughts or behavior. Hence, the purpose of your belief report in **context a** is not to explain Professor Humbert's thoughts or behavior. Rather, your belief report is concerned *de re* with Clare Quilty and his relevance to the history of psychoanalysis. That is why you can be paraphrased as saying "Humbert thinks *of* Clare Quilty that he ...". However, once the ascribee himself and his thoughts are at issue in the conversation (as in **context b**) and, most importantly, once it becomes a goal of the conversation to explain

the ascribee's thoughts and behavior, the belief report is not concerned anymore with Clare Quilty.

So how are you supposed to interpret my utterance of 'Humbert does not believe that Clare Quilty wrote *Psychoanalysis*' in **context b**? The correct interpretation, to my mind, is quite intuitive. What I am saying is not so much that Humbert does not believe of Clare Quilty himself that he authored *Psychoanalysis*, nor am I saying that Professor Humbert does not know who the author of *Psychoanalysis* is. What I am saying is, more likely, something about Humbert's presuppositions concerning the name 'Clare Quilty': e.g., that he does not associate 'Clare Quilty' with the property of authoring *Psychoanalysis*. This naturally prompts the question: how can a speaker use a name 'N' within a belief report in order to convey information about the ascribee's presuppositions about a corresponding referential expression? As I hinted above, I believe this is done through something like the predicate-transfer mechanism described above.

Start with two different and saliently related sets of predicates, P_Φ and P_Ψ, and their corresponding sets of properties Φ and Ψ, defined as:

P_Φ: all predicates $p_{\phi i}$ that apply to N; where 'N' is 'Clare Quilty'.

P_Ψ: all predicates $p_{\psi i}$; where $p_{\psi i} = g_t(p_{\phi i})$; where $g_t(p_{\phi i})$ maps any property $\phi_i \in \Phi$ into the property of naming the bearer of such property—i.e., a corresponding property $\psi_i \in \Psi$.

Φ: every property ϕ_i to which the corresponding predicate $p_{\phi i}$ refers.

Ψ: every property ψ_i to which the corresponding predicate $p_{\psi i}$ refers.

The function $g_t(p_{\phi i})$ is salient: as a matter of fact for every predicate in P_Φ there is a corresponding P_Ψ predicate in virtue of there being a name associated with each ϕ_i property. That there is some such correspondence is simply part of the conversational background. In particular, these sets include two homonymous predicates, $p_{\phi 1}$ and $p_{\psi 1}$, which refer to two different properties, ϕ_1 and ψ_1, respectively.

$p_{\phi 1}$: BEING THE AUTHOR OF *Psychoanalysis*.

$p_{\psi 1}$: BEING THE AUTHOR OF *Psychoanalysis*.

ϕ_1: AUTHORING *Psychoanalysis*.

ψ_1: NAMING THE AUTHOR OF *Psychoanalysis*.

This satisfies the **functionality** criterion on predicate transfer (see page 93). The criterion of **noteworthiness**, as you may recall (see page 93), depends on the conversational interests. It is clear in **context b**—the one exhibiting substitution failure—that talking about names and not their bearers will be useful. In particular, the property of NAMING THE AUTHOR OF *Psychoanalysis* is useful to classify its bearer—i.e., 'Clare Quilty'—relative to the ascribee's mental states—i.e., Humbert's presupposition that 'Clare Quilty' and 'Dr.

Zemph' are *not* coreferential—the representation of which is the goal of the belief report in question. Aside from their intuitive relevance, there are good reasons to think that these presuppositions can be salient in the conversational background. As Swanson 2006 and others have noted, proper names are presupposition-carrying expressions, and these presuppositions may be satisfied by the local context of a belief report—i.e., the ascribee's set of beliefs.[13]

Finally, if by uttering (2a) or (2c) a speaker manages to convey information about the name used in the belief report, it must be fairly obvious, in the conversational background, that in using the name the speaker really intends to mention it.

(2a) Humbert believes that Dr. Zemph wrote *Psychoanalysis.*
(2c) Humbert believes that Clare Quilty wrote *Psychoanalysis.*

In this case what is missing is not an account in terms of meaning transfer or deffered reference but, rather, the simple observation that speakers ordinarily use names in order to mention them. Geurts 1998b has made such observation on independent grounds—i.e., in connection with the phenomena of non-descriptive negations:

> All things considered, language users don't seriously try to make formal distinctions between linguistic expressions and their names. Instead they simply use an expression to refer to itself, whenever the need arises. It follows from this that ALL expressions of any language are equivocal in a way: besides their ordinary meaning(s), they can also be used to designate themselves. [Geurts 1998b, p. 291.]

This explains how in uttering 'Humbert does not believe that Clare Quilty wrote *Psychoanalysis*' in **context b** I let you know something about Humbert's belief states, namely, that he does not associate 'Clare Quilty' with the author of *Psychoanalysis*. The name 'Clare Quilty' is used to refer to itself and the predicate 'being the author . . .' undergoes a transfer of meaning in virtue of which it gets interpreted as referring to a different property: a property that applies to names rather than to bearers thereof. Now, given that the information conveyed is *about the name used*, this also explains why (2a) and (2c) may differ in truth val-

---

[13] Swanson 2006 claims that a speaker's use of a name, say, 'Clare Quilty', presupposes that both speaker and hearer associate the same thing with 'Clare Quilty'. And this presupposition seems to be salient since, according to Swanson, it becomes part of the information that a speaker "gleans at" when being informed about the truth of (1a)-(1b), which is meant to offer an explanation of the difference in informativeness between both sentences. According to this view, what a hearer learns from an utterance of (1b) is that the man she associates with 'Clare Quilty' wrote *Psychoanalysis*.

(1a) Dr. Zemph wrote *Psychoanalysis.*
(1b) Clare Quilty wrote *Psychoanalysis.*

ues. Or, to put it differently, why (2a) and (2b) may both be true in a conversational context where the participants presuppose that 'Dr. Zemph' and 'Clare Quilty' are coreferential:

(2a) Humbert believes that Dr. Zemph wrote *Psychoanalysis.*
(2b) Humbert does not believe that Clare Quilty wrote *Psychoanalysis.*

This is the linguistic account of substitution failure that I favor. It comes with three kinds of virtues. First, it is consistent with the central tenets of a referentialist semantics for referential *uses* of proper names: i.e., that they are singular terms and rigid designators that semantically convey only their referents. Furthermore, the account is consistent with semantic innocence: the idea that one and the same expression has the same meaning inside and outside of 'that...' clauses. This is so because, on my view, when it comes to belief reports with substitution failure the relevant names are used self-referentially and self-referential uses contribute the same meaning (i.e., the name itself) inside or outside of 'that...' clauses.[14] Second, this view is also consistent with Recanati's 1993 constraints (see page 91) on how to solve the underdetermination of 'that...' clauses. Third, by making no use of descriptive information that is linguistically or psychologically associated with proper names, this account is also consistent with the psychological account of substitution failure that I have already argued for (see section 4.3).

4.5 Integrating the accounts

I have offered what I take to be two sides of a single account of substitution failure: the integrative theory. On the psychological side (see section 4.3), I have shown how a subject's cognitive state looks when she fails to realize that two names are coreferential (e.g., the case of Humbert with 'Dr. Zemph' and 'Clare Quilty'). Unless there is enough evidence to the contrary, normal speakers will take different names (e.g., 'Dr. Zemph' and 'Clare Quilty') to name different objects and, hence, they will store the information associated with each name under separate files. Given this localization of information, the subject may have incompatible beliefs about what, as a matter of fact, turns out to be one and the same object.

If the psychological account is credible, then there must be a way in which speakers exploit the cognitive state of the ascribee for the purposes of the conversation. On the linguistic side, the account explains how speakers may use sentences like (2a)-(2c) to convey such information, which results in belief reports that use of different coreferential names

---

[14]Other philosophers who take themselves to be referentialists (e.g., Soames 2002, and Salmon 1986) seem to have a different view. Swanson, in conversation, has said that on this other view names always convey the same meaning. Thus, names always contribute their referent, even in belief reports that exhibit substitution failure. I will not argue here against their view. I am satisfied if my account is considered to be an alternative referentialist theory of proper names.

and have different truth-values.

  (2a) Humbert believes that Dr. Zemph wrote *Psychoanalysis.*
  (2c) Humbert believes that Clare Quilty wrote *Psychoanalysis.*

    To explain this I appeal to the mechanism of meaning transfer for predicates *and* to a self-referential use of proper names. This latter aspect of the account should be underscored. It is not as if, on my view, proper names have two different semantic contents: their referent and themselves. What I am claiming is, rather, that when engaged in the job of ascribing attitudes with substitution failure speakers *use* the relevant names in a self-referential manner and not in the ordinary way. They do so in a way that hearers may *pragmatically* identify that something like this is the right interpretation. It is because of this that the integrative theory retains semantic innocence together with the claims that proper names are singular terms and rigid designators.

    On this view, a belief report like (2b) is properly interpreted as (2d)

  (2b) Humbert does not believe that Clare Quilty wrote *Psychoanalysis.*
  (2d) Humbert does not believe that 'Clare Quilty' names the author of *Psychoanalysis.*[15]

    To see how the integrative theory works it will be useful to consider two different yet closely related long standing philosophical puzzles: Frege's and Kripke's.

4.5.1   Solving Frege's puzzle

The integrated account of substitution failure offers itself as an account of differences in informativeness. To illustrate, consider the following passage from Nabokov's script for Kubrick's "Lolita":

  H: Who is the man that I'm looking for?
  L: Do you remember Dr. Zemph?
  H: Dr. Zemph? Was it him?
  L: Not exactly. Do you remember the car that used to follow us around? Do you remember the guy who called you at the motel?
  H: Yes, I remember very well.
  L: And yet, you still haven't guessed?
  H: Tell me who it was.
  L: It was Clare Quilty.
  H: Who was Clare Quilty?

---

[15]This analysis raises the question as to whether one can say things like 'Humbert believes that Clare Quilty wrote *Psychoanalysis* and consists of two words' to mean something like "Humbert believes that 'Clare Quilty' names the author of *Psychoanalysis* and consists of two words". I address this same worry, with a different example, in section 4.6.7.

L: All of them, of course.

H: You mean, Dr. Zemph, he was Clare Quilty?

L: Yes, Dr. Zemph is Clare Quilty. That's the man you're looking for.

If you have not seen the movie yet (or read the book) you should know that H (i.e., Professor Humbert) has personally met with Dr. Zemph as well as with Clare Quilty. He knows them both by acquaintance. The dialogue shows that Humbert is unable to properly associate the information he is given by L (i.e., Lolita). He remembers pretty well who called him, who he met with at the house (i.e., Dr. Zemph) and who he met with at the motel (i.e., Claire Quilty). Yet he is unable to realize that he has met with the same individual all along. Furthermore, even after he is told that Clare Quilty is the bearer of all these properties, Humbert is still clueless. The end of the dialogue confirms this by pointing out, among other things, that Humbert takes the identity sentence 'Dr. Zemph is Clare Quilty' to be of great informative value.

Even though he is personally acquainted with the referent of both 'Dr. Zemph' and 'Clare Quilty', Humbert, an intelligent professor and writer, has not been given enough evidence to realize that 'Dr. Zemph' and 'Clare Quilty' *do not* denote different objects. Given this absence of contrary evidence, and the fact that 'Dr. Zemph' and 'Clare Quilty' differ phonemically (and orthographically), it is plausible to think (see section 4.2.2) that Humbert takes them to correspond to different objects. Thus, it is natural to think that he stored the names, together with the information associated with them, in different storage spaces. This explains why he remains clueless all through the dialogue, until he is explicitly told that Dr. Zemph is Clare Quilty. On this account, the fact that Humbert finds such a statement to be informative is explained in terms of the information he gleans at from the context together with Lolita's statement.

Lolita knows that 'Dr. Zemph' and 'Clare Quilty' are coreferential but, through the dialogue, she realizes that Humbert presupposes that they refer to distinct individuals. This information, which describes the cognitive state that yields substitution failure according to the psychological account I gave (see section 4.3.1), becomes salient in the context. Lolita can exploit that information in order to convey something about the names used and how they relate to each other. The intuitive interpretation of Lolita's utterance of 'Dr. Zemph is Clare Quilty' is *not* that "the man at the motel is the man in the car". This kind of information has already been presented by Lolita and Humbert has failed to draw the right inferences. The intuitive interpretation is, rather, something like "Dr. Zemph is also named 'Clare Quilty' ." And all this can be done through the same linguistic mechanism that, on the view I defend, speakers exploit in contexts with substitution failure.

This account of Frege's puzzle takes informative identity statements to be properly interpreted in a metalinguistic manner. Thus, some kind of metalinguistic or symbolic understanding is required, on this view, to understand such statements. Perner, Rendl and Garnham 2007 offer some independent psychological evidence supporting this claim. According to Perner and colleagues, a proper understanding of identity statements is tied to an understanding of false belief, both of which seemed to be owed to the development of higher order cognitive abilities.

4.5.2   Solving Kripke's puzzle

According to Kripke 1979 there's something puzzling about belief.

> Peter may learn the name 'Paderewski' with an identification of the person named as a famous pianist. Naturally, having learned this, Peter will assent to "Paderewski has musical talent", and we can infer — using 'Paderewski', as we usually do, to name the Polish musician and statesman:
>
> (6a) Peter believes that Paderewski had musical talent.
>
> Later, in a different circle, Peter learns of someone called 'Paderewski' who was a Polish nationalist leader and Prime Minister. Peter is skeptical of the musical abilities of politicians. He concludes that probably two people ... were both named 'Paderewski'. [...] Peter assents to, "Paderewski had no musical talent". Should we infer ...:
>
> (6b) Peter believes that Paderewski had no musical talent.
>
> [Kripke 1979, p.265.]

How can anyone be in Peter's situation? The integrated account has a simple answer. First, it must be noted that none of the criteria for name counting (see section 4.3.2) deliver the result that 'Paderewski' and 'Paderewski' are two different names. They don't differ phonetically, referentially, or (one might presume) historically. But these criteria do help us understand what goes on with Peter.

In section 4.2.2 I offered some evidence suggesting that in general speakers assume that to every name corresponds only one referent. This assumption, however, was presented as revisable. Whenever they have enough evidence to the contrary, speakers will take, say, two phonemically identical names to denote different objects and, hence, will store them under unrelated files. The account I am proposing takes Peter to be in such a situation.

There is something strange about Peter's case because he has conflicting evidence: he is acquiring a new name (i.e., 'Paderewski') that he assumes refers to one and only one indi-

vidual. But he is also given extra information. He is first introduced to a use of 'Paderewski' that refers to a famous pianist, then to a use of it that refers to a politician. Peter is convinced that politicians are terrible musicians. For Peter, this is evidence that a politician cannot be a famous pianist and, hence, that the two uses of 'Paderewski' that he has been confronted with *do not* in fact refer to the same thing. Briefly put, Peter falsely believes that he has been confronted with two phonemically identical yet referentially *distinct* names. The integrated theory predicts that in these cases subjects will take such uses of the relevant name to count as distinct names and, hence, will be stored in unrelated files.

That is why Peter may be said to believe both that Paderewski had and lacked musical talent without thereby ascribing irrationality to him. He has a false belief, an important one perhaps, but he is certainly not stupid.

4.5.3 Non-linguistic believers?

The integrative account relies on the assumption that the ascribee has some understanding of reference and proper names. There is evidence that this understanding appears in humans at an early age (6 months) and that it constitutes some kind of prelinguistic preparedness (see García-Ramírez and Shatz forthcoming). Thus, this account does not require a full blown linguistic competence on the side of the ascribee. In this sense, it can be extended to account for belief reports that are concerned with non-linguistic believers, so long as they have an understanding of reference and proper names.

I am skeptical, however, that there are any such cases. Elsewhere (see García-Ramírez forthcoming) I have shown that there is substitution failure only when the ascribee can be said to ignore a relevant identity statement. If so, then substitution failure arises only when the belief reports are concerned with linguistic believers—or, at least, with subjects that understand reference, proper names, *and* identity statements.[16] This seems to be supported by the literature. The traditional cases of substitution failure (e.g., Frege's and Kripke's puzzles) very explicitly presuppose linguistic competence on the side of the ascribee.

---

[16]Perner, Rendl, and Garnham 2007 present evidence that an understanding of identity statements requires higher order cognitive abilities, such as those involved in false-belief tasks. Cognitive psychologists associate these abilities with Theory of Mind and metarepresentational competence (see Leslie 1987). There are good reasons to think that these abilities and the degree to which they appear are unique to humans (see Penn, Holyoak, and Povinelli 2008).

4.6   Explanatory power

I started this paper by claiming that the integrative theory can explain some problematic simple sentences that may appear to have substitution failure; and I have been claiming all along that it can also account for substitution failure involving referential expressions other than proper names (e.g., demonstratives). I want to conclude this paper by devoting this section to substantiating these claims. Since the goal is to show how well the integrative theory accounts for otherwise problematic cases, I will be considering many other allegedly problematic sentences. This should be enough to show that we can make an inference to the best explanation from the evidence on proper names, together with its linguistic side, into the case of demonstratives. The result is an integrated account that is our best option when it comes to understanding the psychology and linguistics of substitution failure.

4.6.1   Simple sentences

Let us first go back to Saul's 1997 examples of simple sentences that, allegedly, carry substitution failure.

(7a) Clark Kent came into the phone booth, and Superman came out.
(7b) Dan dresses like Superman.
(7c) Superman was more successful with women than Clark Kent.

I argued before that if we do want to admit there is substitution failure, we must also accept that 'Clark Kent' and 'Superman' are not being used coreferentially. The problem arises when we recall that whoever utters these sentences may at the same time believe that Superman is Clark Kent. How can we account for such a speaker using these names in (7a)-(7c) with substitution failure?

The integrative theory suggests that we look at the context in which (7a)-(7c) are acceptable and see if there is enough information there to guide the interpretation. Saul 1997 doesn't provide us with such a context, but we can safely assume that of the fictional "Superman" comic-book story. To make a long story short, Clark Kent is a Krypton-born, Kansas-raised, and highly decent man with superhuman abilities that he decides to use to benefit humanity. But Clark has a problem: he wants to continue with his normal human life as much as possible. So he decides to pretend to be someone else when using his superpowers; thus, he creates a character which he will enact every time it is needed.[17]

---

[17]There are multiple "Superman" stories. Some take Superman to pretend to be Clark Kent when willing to hide his superhuman powers. Whether it is Clark Kent or Superman who pretends is immaterial for our purposes. So long as it is clear that this particular pretense is involved, everything will be fine.

This character is called 'Superman' and wears blue lycra and red trunks, with a long red cape and an 'S' shield on his chest. Whenever he stops pretending to be Superman, Clark Kent is dressed like any normal person and carries a normal job as a reporter. It is with this information in the background that (7a)-(7c) are uttered.

It seems clear that in all these cases the names 'Clark Kent' and 'Superman' are not coreferential. There are, to my mind, several moves one can make here. It may be that 'Superman' only refers to object x when he pretends to be a superhero (most of the time this happens when he wears blue and red with a cape) and that 'Clark Kent' refers to that same object when he pretends to be a reporter. If we follow this line, we have to claim that 'Superman' and 'Clark Kent' refer to distinct temporal parts of the same object. Alternatively we may want to claim that 'Clark Kent' refers to object x and that 'Superman' refers to a fictional character created and exclusively personified by x. Or perhaps we want to claim exactly the opposite. All of these offer the correct interpretation (7a)-(7c) while explaining why there is substitution failure. The strange thing, as Saul 1997 points out, is to make sense of this while at the same time admitting that (7d) is true.

 (7d) Clark Kent is Superman.

There are two worries one might have about (7d). First, one might worry that it is an acceptable thing to say in the same context in which (7a)-(7c) are successfully uttered. Second, one might worry that it makes little sense for a speaker to utter (7a)-(7c) while believing something like (7d). The integrative theory has an answer to both worries. On the one hand, it is acceptable to utter something like (7d) in the same context of utterance of (7a)-(7c) *because* by uttering (7d) the speaker will put an end to the pretense necessary to make sense of (7a)-(7c). A successful utterance of (7d) presupposes, of course, that the audience doesn't know that the alleged distinctness between Clark Kent and Superman is nothing but a pretense. Otherwise it would not be informative.

On the other hand, a speaker may honestly utter (7a)-(7c) to convey what she takes to be non-trivial information while at the same time believing in something like (7d) *because* she is asked to pretend, for the purposes of the conversation, that either one of the three options above is the case—that Superman and Clark Kent are different temporal parts of the same object, or that Superman is a fiction personified by Clark Kent, or vice versa. There is certainly no problem with pretending that something is the case even though one pretty well knows that it isn't.

Before leaving the field of simple sentences, let me consider a more traditional example owed to Quine 1961. According to this story, Giorgione is Barbarelli, however, (7e) is true while (7f) is false.

(7e) Giorgione was so called because of his size.
(7f) Barbarelli was so called because of his size.

This example has been discussed *ad nauseam* (see Predelli 2009 for a recent descussion). I will not say anything new about it, but will simply point out how the context helps the speaker use 'Giorgione' in a way that exhibits substitution failure. It seems obvious, and has been pointed out by many (see Crimmins 1992), that the use of the predicate 'SO CALLED' makes anaphoric reference to a previously used name. The context is such that the most reasonable way to bind this anaphor is to the name 'Giorgione'. That this is so appears to be verified by the conversation in (7g) which offers a more explicit way of conveying the same information.

(7g) L: Did you hear Barbarelli was called 'Giorgione'?
     E: I didn't know! How come?
     L: He was so called because of his size.

It seems, then, that whoever utters (7e) or (7f) will be using the relevant name both to convey its referent *and* to convey the name itself. There is substitution failure, if at all, precisely because the name is so used. Even if they are coreferential, 'Giorgione' and 'Barbarelli' are phonemically distinct and this difference is exploited in (7e).

This account of (7e) fits well within the lines of the integrative account I am offering. It involves a self-referential use of the relevant name that is purported to not be substitutable, and the predicate 'BEING SO CALLED BECAUSE OF HIS SIZE' is very obviously mapped on to the property of BEING CALLED 'GIORGIONE' BECAUSE OF ONE'S SIZE. Speakers are best interpreted, both here and in the traditional cases (i.e., substitution failure within belief reports), as using the names self-referentially. This is what the integrative theory predicts.

4.6.2   Demonstratives

Substitution failure does not only appear with the use of proper names. Speakers may also use demonstratives in ways that exhibit substitution failure. Consider the following example owed to Clapp (*pers. comm.*).

> Suppose we are hiking with Pierre and we are lost. A disagreement breaks out about where we are, and we begin arguing while pointing at landmarks and positions on various maps. Pierre does not speak English, and you do not speak French. So, in an attempt to make Pierre's views clear to you, I utter:
>
> (8a) Pierre thinks we are *here* but he doesn't think we're *there*.
>
> My utterance might be true, even if the same location is referred to by my uses of 'here' and 'there'.

How can we come up with a metalinguistic interpretation of (8a) and what would that look like? Here is how the integrative theory illuminates this case. It should be clear, first, that (8a) cannot be read *de re*. Suppose the position I point at while uttering (8a) is called 'Point A' on the relevant map (it is a simple-minded map). If (8a) could be read *de re* (8b) should, at least, make sense, but it doesn't.

(8b) Pierre thinks we are at Point A but he doesn't think we're at Point A.

So it must be that, in uttering (8a), I'm exploiting the fact that 'here' and 'there' are phonetically distinct expressions that can nevertheless be used in this context to refer to the same location. Another piece of evidence supporting this is the fact that 'here' and 'there', as they are used in (8a), *must* be phonetically marked. For consider a non-marked use of the same sentence, as in (8c), while you are demonstrably pointing at the same spot on the map.

(8c) Pierre thinks we are here but he doesn't think we're there.

Your belief report would be naturally understood as a joke. Now, the requirement that the demonstratives used be phonetically marked is, perhaps, the clearest evidence that a metalinguistic interpretation is required (see Horn 1989), as evidenced by cases like (8d).

(8d) He looks like Frankenst[á^y]n not like Frankenst[í^y]n.
    That is not a to[ma:]to it's a to[meɪ]to.

All these aspects of the use of (8a) in the mentioned context yield an answer to our first question: a metalinguistic interpretation is triggered by the fact that a *de re* interpretation would not even make sense *and* by the evident fact that the use of the demonstratives is phonetically marked. So what would that metalinguistic interpretation look like? Here's one candidate:

(8e) Pierre thinks 'here' refers to our location but he doesn't think 'there' refers to our location.

There are a couple worries some might have with this interpretation. First of all, it seems strange to say that a monolingual French speaker would have thoughts about the English words 'here' and 'there'. This can be explained by recalling the correspondence principle (see page 88). What the ascriber is trying to do by uttering (8a) is not to be interpreted as literally claiming that Pierre has thoughts about this or that English word, but rather that he has thoughts about the corresponding referential expressions in Pierre's language. Second, it seems strange to say that 'here' or 'there' refer. Isn't it more like it is uses of 'here' and 'there' that refer?

I think there's little to be said about this worry. It seems to me that an adequate interpretation will be one that fits the speaker's goals in the conversation relative to the common

ground. Ordinary speakers would not distinguish between 'here' and uses of 'here'. That, I think, is a well known fact. The distinction is, it seems to me, a theoretical one. So I doubt that we have any reasons to think that saying that 'here' and 'there' refer (label, point, denote, or whatever) will be strange for the participants in the conversation. True, it may seem strange to philosophers of language and linguists, and it may very well be that this conversation takes place among such individuals. In that case I think the claim that 'here' refers may very well seem strange. But, and this is the beauty of the integrative theory, in such case we would have salient presuppositions and functions allowing us to get a different interpretation, something like (8f).

(8f) Pierre thinks 'here' is used to refer to our location but he doesn't think 'there' is used to refers to our location.

How could we get such a theory-laden interpretation? Well, it's quite simple. If the participants are theorists that find it strange to claim that 'here' refers, it must be because they strongly believe in some token/type distinction and that it's only *uses* of referential expressions that manage to refer. The integrative theory claims that the relevant meaning-transfers will depend always on the context and the functions that are made salient by it. The theoretical beliefs of the participants seem to be quite salient. So I do not see why it cannot deliver a function mapping the predicate 'BEING HERE' onto the property of BEING AT A LOCATION REFERED TO BY THIS USE OF 'HERE'.

4.6.3 Anaphoric reports

Substitution failure can also appear within externally bound anaphoric constructions as in (9a).[18]

(9a) Superman came in but Lois didn't realize he was Clark Kent.

Unlike previous cases, the pronoun 'he' is anaphorically bound by the referential use of 'Superman' in the sentence context. So we seem to be forced to interpret 'he' as referring to Superman. That is fine with me, so long as the context is sensitive to the use of the name 'Clark Kent' such that BEING NAMED 'CLARK KENT' is a noteworthy property. In such a case there will be a function mapping the property of BEING CLARK KENT (if there is some such property) into the property of BEING NAMED 'CLARK KENT'. If (9a) genuinely exhibits substitution failure, then the context will provide the relevant properties and functions.

---

[18]Eric Swanson mentioned this example in conversation. Similar examples appear already in Crimmins 1998.

So, according to my view, (9a) is properly interpreted as (9b).

(9b) Superman came in but Lois didn't realize that he was named 'Clark Kent'.

This, to my mind, seems like a very reasonable thing to say in such a context. All is fine with externally bound anaphoric belief report constructions. This same story accounts for another related example as in (9c):

(9c) Superman came in but Lois didn't realize he was Clark Kent, and in virtue of that identical to Kal-El.

Suppose the context is the same as that of (9a). If so, then BEING NAMED 'KAL-EL' is a noteworthy property in that context. Hence, there will be a function mapping the property of BEING IDENTICAL TO KAL-EL (if there is some such property) into the property of BEING NAMED 'KAL-EL'. If (9c) *genuinely exhibits substitution failure*, then the context will provide the relevant properties and functions. So, according to my view, (9c) is properly interpreted as (9d):

(9d) Superman came in but Lois didn't realize that he was named 'Clark Kent' and, in virtue of that, that he was also named 'Kal-El'.

It may seem odd to say that someone is named 'X' in virtue of being named 'Y'. This oddity is, on this view, contextually explained. If the context is such that, say, *by law* all individuals named 'Alexander' must also be named 'Jon', then it will be acceptable to say something like (9e):

(9e) He is named 'Alexander' and, in virtue of that, he is also named 'Jon'.

When the proposition *that all the relevant individuals named 'X' are named 'Y'* is presupposed in the context, it seems fine to utter 'He is named 'X' and, in virtue of that, he is also named 'Y' '. Thus, the integrative theory takes (9c) to be acceptable in a context where it is presupposed *that there is one individual named 'Clark Kent'* and *that the individual named 'Clark Kent' is the only one named 'Kal-El'*. In such a context, uttering either (9c) or (9d) would be felicitous.

(9c) Superman came in but Lois didn't realize he was Clark Kent, and in virtue of that identical to Kal-El.
(9d) Superman came in but Lois didn't realize that he was named 'Clark Kent' and, in virtue of that, that he was also named 'Kal-El'.

To contrast, consider a context that does not presuppose *that the individual named 'Clark Kent' is also named 'Kal-El'* or *that there is only one person named 'Kal-El'*. As far as we know there might be several people named 'Kal-El' and we are looking for one of them. A man comes in wearing a name tag that reads 'Clark Kent'. In such a context, it would be

infelicitous for a third party to utter (9f) or (9g).

(9f) Did you see him? He was Clark Kent, and in virtue of that identical to Kal-El.

(9f) Did you see him? He was named 'Clark Kent', and in virtue of that he was also named 'Kal-El'.

We would be justified to ask: what do you mean "in virtue of that"?

4.6.4 Mistaken identities

Here's another case owed to Crimmins 1998 (and apparently also to Steve Yablo). As we very well know, speakers may be mistaken about certain identity statements. Andy, for example, may very well not know that Dr. Zemph is Clare Quilty. Suppose then that Andy takes these names to have different referents. How should we understand (10a) as it is used by Andy?

(10a) Humbert believes that Dr. Zemph is smarter than Clare Quilty.

On the view I have argued for, what Andy does not know is that 'Dr. Zemph' and 'Clare Quilty' are two names of the same person. He certainly does not want to convey the information that Professor Humbert believes that Dr. Zemph is smarter than himself or than some merely possible individual. It is at least unclear how to make sense of Andy's utterance. On my view, the best way to understand (10a), after applying the proper meaning-transfer operations, is to interpret it either as (10b) or as (10c).

(10b) Humbert believes that 'Dr. Zemph' names someone who is smarter than Clare Quilty.

(10c) Humbert believes that 'Dr. Zemph' names someone who is smarter than the person named 'Clare Quilty'.

I think both interpretations are good enough for the purposes at hand, and both are suggested by the integrative theory.

4.6.5 Embedded reports

Attitude reports may (apparently) also be embedded within attitude reports. Here is an example owed to Crimmins 1998.

(11a) Andy believes that Humbert doubts that Clare Quilty wrote *Psychoanalysis*.

How should we make sense of this? To find a plausible candidate for the content of an utterance of (11a) we need a lot more contextual information. Such information will help us understand whether the important information concerns Andy's mental states or

Humbert's. Furthermore, without this information, we cannot know whether (11a) is subject to substitution failure at any level of embedding. In any case, assuming there is substitution failure at any of the embedded sentences, the mechanism of predicate meaning-transfer will help us find a plausible candidate for the content of the assertion (or the ascribed mental state). There is nothing precluding predicate transfer from targeting the most or the least embedded sentence or any combination thereof. Just as the simple version of predicate transfer applies to ordinary first-order attitude reports, the same account can be applied to less ordinary second (or higher?) order attitude reports. Like predicates, predicate transfer can be embedded.

4.6.6   Anaphoric and *de se* reports

Other problem cases include what has come to be known as *de se* ascriptions. Once more, the examples come from Crimmins 1998.

(12a) Sometimes, when I see an important philosopher, I don't realize that she *is* that philosopher [or: that she is *she*].
(12b) That man doesn't think that he is Lingens.
(12c) I think that he is Lingens.
(12d) Unlike you, I think that you are Lingens.

   I have already shown how to deal with externally bound anaphoric belief-report constructions. So called *de se* ascriptions are not too different. The suggestion before consisted in taking the predicate 'BEING X', where 'X' is a referential term, as transferring its meaning from the property of BEING X (if there is such property) on to the property of BEING NAMED 'X'. In the present case all we need is for the predicate in question (e.g. 'BEING THAT PHILOSOPHER') to transfer its meaning on to a suitably different property: i.e., that of BEING NAMED AS X or, perhaps, REALLY BEING NAMED 'X'. This strategy delivers the following interpretations of (12a)-(12d).

(12e) Sometimes, when I see an important philosopher, I don't realize that she *is named as* that philosopher [or: that she *is named as she is*].
(12f) That man doesn't think that he *is really named 'Lingens'*.
(12g) I think that *his real name is 'Lingens'*.
(12h) Unlike you, I think that *your real name is 'Lingens'*.

   If the contexts of utterance of (12a)-(12d) do in fact exhibit substitution failure, then the names used will be salient and the property of having those particular names will very likely be noteworthy. Suppose for example that you are going through passport control and the officer wonders whether you are the infamous Nazi official Rudolf Lingens. On his

view, you have a fake passport and a modified facial expression. You insist that you are, say, Gottlob. He replies by uttering (12d): Unlike you, I think you are Lingens. In such a case, I believe, (12h) offers a very plausible interpretation of the corresponding utterance (or mental state). Similar remarks apply to (12e)-(12g).

4.6.7 Germans

Last, but not least, something must be said about the limits of the account. My theory borrows heavily from Nunberg's 1979 and 2004 account of predicate meaning-transfer. Now, according to this theory, we can say things like (13a).

(13a) Edu's dissertation, which weighs five pounds, has been refuted.

whereby we manage to say that an abstract thing, which has a physical instance that weighs five pounds, has been refuted. According to Nunberg, we do this by mapping the predicate 'WEIGHING FIVE POUNDS' into the property of HAVING A CONCRETE INSTANCE THAT WEIGHS FIVE POUNDS. It seems, then, that speakers can transfer meanings to almost any property. In particular, my theory seems to predict that we can say things like (13b)[19]

(13b) Humbert believes that Dr. Zemph, which is a German name, is a central figure of Psychoanalysis.

But, as Recanati has pointed out to me, (13b) is not really an acceptable thing to say. So, it seems, my theory makes a wrong prediction, unless it has a good explanation of why we cannot say things like (13b).

Fortunately, there is some such explanation. It is not true, as it may seem at first glance, that any predicate can transfer its meaning to any alternative property; even predicate transfer has its limits. Consider, for example, a conversation among contemporary artists. Nunberg 2004 notes that in such a context it is acceptable to say things like (13c) but not things like (13d).

(13c) I'm in the *Withney*.
(13d) I'm in the second hall, first cradle to the left.

What is the difference between (13c) and (13d)? Why can we say one but not the other? Nunberg has a simple explanation, which I find convincing: it is because the property of HAVING YOUR WORK EXHIBITED AT THE *Withney* is a noteworthy property in the context, while that of HAVING YOUR WORK EXHIBITED IN THE SECOND HALL is not. The same, I think, goes on with (13b).

---

[19]It is important to note that this same result appears for "ham sandwich" cases. This suggests that this is a more general problem and not a unique feature of my account.

(13b) Humbert believes that Dr. Zemph, which is a German name, is a central figure of Psychoanalysis.

We cannot say things like (13b) because, unlike the property of BEING NAMED 'DR. ZEMPH', the property of HAVING A GERMANIC NAME is not a noteworthy property in contexts with substitution failure. It's not true, strictly speaking, that the integrative theory predicts that we can felicitously utter (13b) *simpliciter*. What it really predicts is that in contexts where the property of HAVING A GERMANIC NAME is noteworthy, it will be felicitous to utter something like (13b). I think this is true. To illustrate, consider **context c**.

context c:

> Andy and I are discussing the history of psychoanalysis. He claims that it is mainly a product of French culture, while I think it has a strong German influence. To determine who is correct we decide to look at the "big names" of psychoanalysis and consider whether they are German or French. To defend his point, Andy goes on to mention the names of some well known French psychoanalysts: "Look, Lacan, Dolto, Roudinesco, Lagache, Favez-Boutonnier, and Reverchon-Jouve—they are all French names." Knowing how highly Andy thinks of Professor Humbert's knowledge of the history of psychoanalysis, I reply by saying: "Yes, but Humbert believes that Dr. Zemph, which is a German name, is a central figure of psychoanalysis."

This, to my mind, is a perfectly acceptable use of (13b).

References

Anderson, J., and Hastie, R. 1974: "Individuation and reference in memory: proper names and definite descriptions". *Cognitive Psychology* 6, 495-514.

Avila, C., Lambon Ralph, M.A., Parcet, M.A., Geffner, D., and Gonzalez-Darder, J.M. 2001: "Implicit word cues facilitate impaired naming performance: evidence from a case of anomia". *Brain and Language* 79, 2, 185-200.

Bloom, P., 1990: "Syntactic distinctions in child language". *Journal of Child Language*, 17, 343-55.

Cohen, G. 1990: "Why is it difficult to put names to faces". *British Journal of Psychology* 81, 287-97.

Crimmins, M. and Perry, J., 1989: "The prince and the phone booth: Reporting puzzling beliefs". *The Journal of Philosophy* 86, 12, 685-711.

Crimmins, M., 1992: *Talk About Beliefs*. Cambridge, MA: MIT Press.

Crimmins, M., 1998: "Hesperus and Phosphorus: Sense, pretense and reference". *Philosophical Review* 107, 1, 1-47.

Diaz, F., Lindin, M., Galdo-Alvarez, S., Facal, D., Juncos-Rabadán, O. 2007: "An event-related potentials study of face identification and naming: the tip-of-the-tongue state". *Psychophysiology* 44, 50-68.

Donellan, K., 1966: "Reference and definite descriptions". *The Philosophical Review* 75, 281-304.

Donellan, K., 1970: "Proper names and identifying descriptions". *Synthese* 12, 335-58.

Frege, G., 1892: "On sense and nominatum". In A.P. Martinich, (ed.) 2002: *The Philosophy of Language*. Oxford: Oxford University Press.

Fodor, J., 1990: "Substitution arguments and the individuation of belief". In J. Fodor, 1990: *A Theory of Content*. Cambridge, Mass: MIT Press.

García-Ramírez, E., and Shatz, M., forthcoming: "On problems with descriptivism: Psychological assumptions and empirical evidence". *Mind & Language*.

García-Ramírez, E., forthcoming: "The puzzle of names". Typescript. Ann Arbor: University of Michigan.

Gelman, S. A., and Taylor, M., 1984: "How two-year-old children interpret proper and common names for unfamiliar objects". *Child Development* 55, 1535-40.

Geurts, B., 1998a: "Presupposition and anaphors in attitude contexts". *Linguistics and Philosophy* vol.21, 545-601.

Geurts, B., 1998b: "The mechanisms of denial". *Language* 74, 2, 274-307.

Goodglass, H., and Wingfield, A. 1993: "Selective preservation of a lexical category in aphasia: dissociations in comprehension of body parts and geographical place names following focal brain lesion". *Memory* 1, 4, 313-28.

Hadar, U., Jones, C., and Mate-Kole, C. 1987: "The disconnection in anomic aphasia between semantic and phonological lexicons". *Cortex* 23, 3, 505-17.

Hall, G. 1996: "Preschoolers' default assumptions about word meaning: Proper names designate unique individuals". *Developmental Psychology* Vol. 32, No.1, 177-86.

Hall, D.G. 2002: "Semantics and the acquisition of proper names" In R. Jackendoff, P. Bloom, and K. Wynn, (eds) 2002: *Language, Logic and Concepts*. Cambridge: MIT Press.

Hall, G., and Graham, S., 1999: "Lexical form class information guides word-to-object mapping in preschoolers". *Child Development* Vol. 70, No.1, 78-91.

Heim, I., 1983: "On the projection problem for presuppositions". In P. Portner and B. H. Partee, (eds.) 2002: *Formal Semantics: The Essential Readings*. Oxford: Blackwell.

Horn, L., 1989: *A Natural History of Negation*. Chicago: University of Chicago Press.

Jacob, P., 1990: "Semantics and psychology: The semantics of belief ascriptions". In N. Cooper and P. Engel (eds.) 1990: *New Inquiries into Meaning and Truth*. Hemel Hempstead: Harvester Wheatsheaf.

Jackson, F. 1998a: "Reference and description revisited". *Philosophical Perspectives* 12, 201-218.

Kambanaros, M. and van Steenbrugge, W. 2006: "Noun and verb processing in Greek-English bilingual individuals with anomic aphasia and the effect of instrumentality and verb-noun name relation". *Brain and Language* 97, 2, 162-177.

Kanehman, D., and Treisman, A., 1993: "Changing views of attention and automaticity" in R. Parasumaran and J. Beatty, (eds.) *Varieties of Attention*. New York: Academic Press.

Kaplan, D., 1989: "Demonstratives". In J. Almong, H. Wettstein, and J. Perry (eds.) 1989: *Themes from Kaplan*. New York: Oxford University Press.

Kaplan, D., 1990: "Words". *Aristotelian Society. Supplementary Volume* LXIV, 93-119.

Katz, N., Baker, E., and Macnamara, J., 1974: "What's in a name? A study of how children learn common and proper names". *Child Development*. 45, 469-73.

Kripke, S., 1979: "A puzzle about belief". In A. Margalit, (ed.) 1979: *Meaning and Use*. Dordrecht: Synthese.

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

Le Dorze, G., and Nespoulous, J. 1989: "Anomia in moderate aphasia: problems in accessing the lexical representation". *Brain and Language* 37, 3, 381-400.

Leslie, S.J., 2008: "Generics: Cognition and acquisition". *Philosophical Review* 117, 1, 1-47.

Lucchelli, F. and Renzi, E. 1992: "Proper name anomia". *Cortex* 28, 221-230.

Macnamara, J., 1982: *Names for Things. A Study of Human Learning*. Cambridge, MA: MIT Press.

Marcus, R.B., 1961: "Modalities and intensional languages". *Synthese* 13,4, 303-22.

Nunberg, G., 1979: "The non-uniqueness of semantic solutions: Polysemy". *Linguistics and Philosophy* 3, 143-84.

Nunberg, G., 1993: "Indexicality and deixis". *Linguistics and Philosophy* 16: 1-43.

Nunberg, G., 1995: "Transfers of meaning". *Journal of Semantics* 12, 2, 109-32.

Nunberg, G., 2004: "The pragmatics of deferred interpretation". In L. Horn (ed) 2004: *The Handbook of Pragmatics*. Malden, MA: Blackwell Publishing.

Penn, D.C., Holyoak, K.J., and Povinelli, D.J., 2008: "Darwin's mistake: Explaining the discontinuity between human and nonhuman minds." *Behavioral and Brain Sciences*, 31, 109-178. Cambridge: Cambridge University Press.

Perner, J., Rendl, B., and Garnham, A., 2007: "Objects of desire, thought, and reality: Problems of anchoring discourse referents in development." *Mind & Language* 22, 5, 475-513.

Predelli, S., 2009: "Substitutivity, obstinacy, and the case of Giorgione" in *Journal of Philosophical Logic* available online: http://www.springer link.com/ content/w3621897514441kx/.

Putnam, H., 1975: "The meaning of 'meaning' ". In H. Putnam, 1975: *Mind, Language and Reality: Philosophical Papers*. Vol. 2. Cambridge: Cambridge University Press.

Quine, W.V.O., 1961: "Reference and modality". In W.V.O., Quine, 1961: *From a Logical Point of View*. New York: Harper and Row.

Recanati, F., 1993: *Direct Reference*. Oxford: Blackwell.

Recanati, F., forthcoming: "Singular thought: In defence of acquaintance". In R. Jeshion (ed.) forthcoming: *New Essays on Singular Thought* Oxford: Oxford University Press.

Richard, M., 1990: *Prospositional Attitudes*. Cambridge: Cambridge University Press.

Salmon, N., 1986: *Frege's Puzzle*. California: Ridgeview Publishing.

Salmon, N., 1989: "Illogical belief". *Philosophical Perspectives Vol. 3, Philosophy of Mind and Action Theory* 243-85.

Saul, J.M., 1997: "Substitution and simple sentences". *Analysis* 57, 2, 102-8.

Schiffer, S., 1978: "The basis of reference". *Erkenntnis*, 13, 171-206.

Semenza, C., 2009: "The neuropsychology of proper names". *Mind & Language* 24, 4, 349-71.

Semenza, C., and Sgaramella, T.M. 1993: "Production of proper names: a clinical study of the effects of phonemic cueing". *Memory* 1, 4, 265-80.

Semenza, C., and Zettin, M. 1989: "Evidence from aphasia for the role of proper names as pure referring expressions". *Nature* 342, 678-79.

Soames, S., 1989: "Direct reference, propositional attitudes, and semantic content". *Philosophical Topics* 15, 44-87.

Soames, S., 2002: *Beyond Rigidity*. Oxford: Oxford University Press.

Sommer, W., Komoss, E., and Schweinberger, S.R., 1997: " Different localization of brain systems subserving memory for names and faces in normal subjects with event-related potentials". *Electroencephalography and Clinical Neurophysiology* 102, 192-99.

Stalnaker, R., 1978: "Assertion". In R. Stalnaker, 1999: *Context and Content*. Oxford: Oxford University Press.

Stalnaker, R., 1987: "Semantics for belief". In R. Stalnaker, 1999: *Context and Content*. Oxford: Oxford University Press.

Stalnaker, R., 1988: "Belief attribution and context". In R. Stalnaker, 1999: *Context and Content*. Oxford: Oxford University Press.

Swanson, E., 2006: "Frege's puzzle and pragmatic presupposition". In E. Swanson, 2006: *Interactions with Context.* PhD dissertation. MIT.

Treisman, A., 1992: "L'atention, les traits, et la perception des objets". In D. Andler, (ed.), 1992: *Introduction aux Sciences Cognitives*. Paris: Gallimard.

Valentine, T., Brennen, T., and Brédart, S. 1996: *The Cognitive Psychology of Proper Names*. London: Routledge.

Ward, G., 2004: "Equatives and deferred reference". *Language* 80, 2, 262-89.

Werheid K., and Clare L., 2007: "Are faces special in Alzheimer's disease? Cognitive conceptualisation, neural correlates, and diagnostic relevance of impaired memory for faces and names". *Cortex* 43, 898-906.

Chapter 5

The Cognitive Theory of Empty Names

Ordinary use of empty names encompasses a variety of different phenomena, including issues in semantics, mental content, fiction, pretense, and linguistic practice. In this paper I offer a novel account of empty names, the cognitive theory, and show how it offers a more satisfactory account of the phenomena than a recent account owed to Sainsbury 2005 and 2010. The paper includes a brief discussion of issues in metaphysics and psychological explanation, and concludes with a metalinguistic account of negative existentials.

Introduction

My goal is to offer an account of empty names that does not take them to have descriptive semantic content. Doing so will allow me to keep a homogenous semantics for referential uses of names (empty or otherwise) that is consistent with the empirical evidence (see García-Ramírez and Shatz forthcoming). In this sense, I take my view to be anti-descriptivist. There are two different kinds of anti-descriptivist theories of empty names in the literature. On the one hand, there are those (see McDowell 1977, 1984, and 1986, and Evans 1981, and 1982) that claim that the actual reference of names is an essential part of content and, hence, that failures of reference result in failures of contentful thinking. On this view, empty names are not intelligible. I will not be following this line of thought. As I will show later on, there are good reasons to think empty names may be used to express intelligible, full blown thoughts.

On the other hand, there are those that agree that empty names are intelligible, but differ with respect to the kinds of content they take empty names to contribute. Within this group, there are theories of two different sorts: complete and incomplete contents. Braun 1993 and 2005, Reimer 2001a and 2001b, and Taylor 2000, seem to think that empty names are used to express either "gappy propositions" or "propositions in waiting". These views either take the content to be pragmatically completed by descriptive information derived from

the context, or take speakers to always assert falsehoods by using empty names. I take the relevant content associated with empty names to be complete, non-descriptive, and subject to be true. So I will not be following this strategy.

Two kinds of complete contents have been associated with empty names: indefinite propositions such as *that something has property* F (see Adams and Stecker 1994) and singular propositions such as *that A is F* (see Salmon 1998, Soames 2002, and Thomasson 2003; see Sainsbury 2010 for objections). My account is, in a sense, reminiscent of the latter view. I do think that, in the relevant contexts, empty names are used to express complete contents that resemble singular propositions. Yet my view is importantly distinct. On the opposing view, empty names are *not* empty: speakers manage to express singular propositions by using them *because* the names do in fact have a referent. Thus, speakers manage to express complete singular propositions by using empty names *in virtue of* the semantics of the name.

I do not agree with the idea that empty names have referents. It requires that we accept an unnecessary ontology of fictional and other abstract objects. On my view empty names are empty and speakers manage to express complete contents *in virtue of* general cognitive mechanisms of interpretation. Thus, on my view, we need more than just the semantics of names to understand what goes on with the ordinary use of empty names. Part of the goal of this paper is to describe such machinery.

In section 5.1 I present a survey of the phenomena involved in the ordinary use of empty names. There I show how Sainsbury's 2005 RWR-theory accounts for the phenomena, and offer two objections against it. Section 5.2 presents the central claims of the cognitive theory I wish to defend. I develop the theory by showing how it solves the puzzles in section 5.3. Section 5.4 presents the advantages of the theory by showing how it can be used to address further worries about fictional entities. I conclude in section 5.5 by offering an account of the use of empty names in negative existential constructions.

5.1 A theory and the phenomena

Reimer 2001a identifies two problems related to the ordinary use of empty names. On the one hand, speaker intuitions support the idea that statements using empty names are meaningful and truth-evaluable. Reimer dubs this "the intuition problem". On the other hand, speakers' behavior suggests that there is something which they affirmed, denied, etc. while using empty names. As theorists, we are left with the task of describing some proposition or other that is asserted, denied, etc. Reimer calls this "the proposition problem".

I believe there is more to this story. Speaker behavior points to many other problematic tasks left for us to work out. Competent speakers do not only use empty names to make simple declarative statements; they also use empty names to engage in games of make-believe, and to make belief reports that describe behavior adequately, both within and without games of make-believe.[1] The use of empty names within games of make-believe demands an account that is consistent with a more general theory of fictional discourse and, as a consequence, of pretense. Adequate belief reports that make use of empty names suggest that the proposition expressed by the embedded *that* clause has all the explanatory virtues of a proposition that would be expressed by using a non-empty name. A satisfactory account of empty names must not only solve the "intuition" and "proposition" problems, it must do so in a way that is fitting for attitude ascriptions and fictional uses of empty names.

Sainsbury 2005 and 2010 presents a theory meant to solve these problems by defending the following central claims:

RWR 1: All referential expressions, and particularly names (empty or not), are semantically homogenous. They all have the same semantic behavior and determine truth-conditions in the same way.

RWR 2: All names are associated with referential conditions of the form:
$\forall x$ ('N'$_R$ x $\equiv$ x=N)
where $X_R Y$ takes X to refer to Y.

RWR 3: 'S is p' is false if and only if either (i) 'S' lacks a referent; or (ii) 'S' 's referent lacks the property referred to by 'p'.

Together RWR 1 and RWR 2 entail that all names, empty or not, have a meaning: i.e., the referential condition described by RWR 2. There can be, on this view, reference without referents (hence the label 'RWR'). Non-empty names will have a referent on top of this. This allows us to understand the intelligibility of empty names. According to this view, an assertion of 'N is F', where 'N' is an empty name and 'F' is a predicate, will express something like:

$\forall x$ ('N'$_R$ x $\equiv$ x=N $\wedge$ Fx).

This solves the proposition problem, mentioned by Reimer. The content of a sentence using empty names is modeled by the quantified first order logic sentence above. According to RWR 3 all affirmative sentences using empty names are false. So this solves the truth-value

---

[1]See section 5.5 on uses of empty names in negative existential claims.

problem: all assertions using empty names are truth-evaluable. All affirmative ones are, in fact, false. Sainsbury thinks this also delivers an account of how negative existential assertions may express truths. For 'S doesn't exist' just is the negation of 'S exists'. If 'S' is empty, then 'S exists' is false and, hence, its negation (i.e., 'S doesn't exist') is true.

I have two objections against RWR, which I will present in the following sections. For now let me just point out that RWR does not really treat empty and non-empty names in a homogenous semantic way. On this view, non-empty names always (i.e., by necessity) have a referent, while empty names never (i.e., by necessity) have one. And there is no intermediate referring expression according to the theory. With this and RWR 3 it's not difficult to show that empty and non-empty names always get their truth-conditions through different means. Sentences with non-empty names are always true/false depending on whether the referent has/lacks the relevant property. Sentences with empty names are always false because they lack a referent.

To illustrate suppose we hold fixed an ordinary interpretation of the predicates 'is tall' and 'is short'. If so, then there is no single context in which 'Mark Sainsbury is tall' and 'Mark Sainsbury is short' can both have the same truth-value.[2] And the same happens if we use instead the name 'Barack Obama' or any other non-empty name. This, however, is not true of empty names. According to RWR 3, all affirmative assertions using empty names are false. If so, then there is at least one single context, namely, a non-fictional one, where 'Hamlet is tall' and 'Hamlet is short' both have the same truth-value (i.e., false).

This result seems odd, in and of itself. But the problem is not limited to its oddity. The problem is that it's not clear how we can accept it and still claim that there is no semantic difference between empty and non-empty names, as Sainsbury 2005 wants to claim. RWR does not offer a genuinely homogenous semantic way to treat empty and non-empty names.

5.1.1 Mental content

Sally believes that Santa is coming tonight. She leaves cookies and milk near the Christmas tree and prepares herself to stay up all night. How should we describe Sally's mental state? Is it a typical belief-like propositional attitude? If so, what is the proposition expressed by the embedded clause *that Santa is coming tonight*? As suggested above, according to RWR the embedded sentence expresses something like:

---

[2]I am aware of the fact that 'Mark is tall' and 'Mark is short' can both be used to say something true provided we change the standards of measurement. But there is no way to change the standards of measurement without changing the context. So it's still the case that 'Mark is tall' and 'Mark is short' cannot both be used with the same truth-value *in the same context*.

$$\forall x \, (\text{'SANTA'}_R \, x \equiv x{=}\text{SANTA} \wedge Cx).$$
(where Cx: x is coming tonight)

This account helps illuminate the explanatory role of belief reports because it allows us to distinguish among the content conveyed by using different empty names. Using 'Hamlet' instead of 'Santa' will help us convey different propositions. This is so in virtue of the fact that the reference conditions are, as Sainsbury puts it, homophonic. They involve the names that they define. Thus, if we can distinguish between 'Hamlet' and 'Santa' we can distinguish between claims about Hamlet and claims about Santa, as well as between beliefs about the former and beliefs about the latter.

The belief report 'Sally believes that Santa is coming tonight' is successful, on this view, because it takes the subject to distinguish all the predicates she associates with whichever object meets the condition of being identical with Santa from, say, objects that meet the condition of being identical with Hamlet. Enjoying cookies and milk by the Christmas tree is not a property that competent speakers associate with whichever object is Hamlet. It is, apparently, one they associate with whichever object is Santa.

This gets us to my first objection against RWR. Like any account of the semantics of names, this explanation comes with a cognitive cost: in order to understand names subjects must also understand identity statements of the form 'x=A' or, alternatively, they must understand how an object may have the property of BEING IDENTICAL WITH A. There is substantial empirical data against this view (see García-Ramírez and Shatz forthcoming). On the one hand, there are good reasons to think that 6-month-olds do competently understand their own names, even though they have a poor and general understanding of properties (i.e., they only understand general properties such as HAVING A CERTAIN SHAPE). On the other hand, recent studies (see Perner, Rendl, and Garnham 2007) suggest that identity statements are not understood until between the second and third year of life. Thus, by the time they are 18 months old, human infants already understand several different names, yet they lack an understanding of identity. This strongly suggests, against Sainsbury's account, that identity conditions are not part of the semantics of proper names.

5.1.2 Fictional discourse

Santa is a jolly old man with a white beard who lives in the North Pole and wears red. This is fiction. As a matter of fact, there is no such thing as Santa. So the short story above cannot be true. This naturally prompts the question: what is the difference between fictional and non-fictional discourse? Why is this Santa story so different from saying, for example, that Bush is a jolly old man with a white beard who lives in the North Pole and wears red?

Whichever theory of empty names we may have, it better be consistent with a satisfactory distinction between fictional and non-fictional *discourse*.

Walton (1990) presents a challenge for any such account.[3] There are, at least, two requirements: First, the account must be consistent with a more general distinction between fiction and non-fiction—the distinction between works of art and artifacts, between sculptures and chairs—and not just between fictional and non-fictional *discourse*. Second, it must explain why it is that works of fiction, unlike works of non-fiction, warrant their own assertions. It is *because* the fiction says so that one is warranted to assert that Santa wears red. However, it is not because Darwin says so that one is warranted to assert that species evolve. Here's a better way to put it.

> A particular work of fiction, in its context, establishes its fictional world and generates the fictional truths belonging to it. [. . .] Every piece of discourse or thought which aspires to the truth has a reality independent of itself to answer to [. . .] The fictional world corresponding to a given work of fiction is not thus independent of it. [Walton 1990, p. 101-102.]

As I presented it, RWR claims that, in virtue of semantics alone, all assertions using empty names are false. Thus, the self-warranting aspect of fictional discourse is not accounted for in virtue of the semantics alone. According to Sainsbury 2010, successful assertions of 'Santa is coming' carry a fiction operator and can be either paraphrased or replaced by assertions of 'According to the story, Santa is coming'. The account is sensitive to how fictional discourse warrants its own assertions, since assertions of the latter form can, in virtue of their semantics, be used to express something true. Appealing to fiction operators, however, undermines Sainsbury's account. For it fails to meet the first of Walton's requirements: it cannot be generalized to explain the difference between fiction and non-fiction, between chair-shaped sculptures and chairs.

This is my second objection to RWR. Sainsbury's is, very clearly, a linguistic strategy. As such it should be rejected for the same reasons that Walton 1990 rejects other linguistic accounts such as Searle's 1975 (see also Schiffer 1996): i.e., because they make nonsense of the general fiction / non-fiction distinction. Works of fiction go way beyond discourse. Most of them are not assertions, let alone assertions whose truth value must be evaluated within a story. Are paintings assertions? Do they tell us a story? Are Rothko's *blue-over-red* works of fiction telling us something according to which something else is true? The answer is clearly negative. The fictional / non-fictional *discourse* distinction simply cannot be based upon linguistic features (like having or lacking a fiction operator), for the same distinction must be drawn in non-linguistic contexts. There is no way around this. Walton's 1990 lesson

---

[3]See Walton 1990, especially Chapter 2 "Fiction and nonfiction".

seems as clear as truthful: we must resist the temptation to treat fictionality and fictional truth as a species of truth (e.g., as truth with a fiction operator). If paintings and sculptures are fictions, just like Santa stories, with nothing close to fiction operators, then it cannot be that fiction operators distinguish fictional form non-fictional *discourse*.

If having fiction operators is not what makes fictional discourse fictional, what does? To answer this question I believe we must take the self-warranting aspect of fictional discourse seriously. Here I follow Walton 1990 up to some extent. I believe that what makes a piece of discourse a fictional one is the fact that it prescribes its own interpretation. Whether a given sentence *S* is to be interpreted as meaning that Φ is the case or that Ψ is the case is something that is determined by the fictional story. Non-fictional assertions do not, in this sense, prescribe their own interpretations. Fictional names do not have a semantic value to contribute, for they do not have a referent. They are meaningful mainly in virtue of the relevant fictional context where they appear. Empty names, in general, do not have a semantic value to contribute. They are meaningful mainly in virtue of the relevant context in which they appear.

It should be clear that by 'context' I do not mean anything linguistic (this is not a linguistic strategy). The contexts I have in mind are more cognitively general, representational contexts, which may be understood as situations involving competent subjects and objects taken as representations. Similarly, when I say that fictions prescribe their own interpretations I do not mean to say that they prescribe their own semantic values. The term 'interpretation' here must be understood in a more general sense that includes several kinds: linguistic, aesthetic, moral, etc. If the relevant piece of fiction is linguistic, then the context will probably provide an interpretation of a linguistic kind.

Thus, on my view, the distinction between fictional and non-fictional discourse is non-semantic: for something to count as fictional discourse a speaker must assume that its correct interpretation is given by the fiction itself, and not merely by the semantic or lexical conventions that prevail outside of the act of fiction making. There is no fictional discourse outside of the fictional context (though it may very well be that the same string of symbols is meaningfully used both inside and outside of fictional contexts). And the same goes on with non-linguistic pieces of fiction.

To wrap up, what makes a piece of discourse something fictional is *not* the fact that it lacks any meaning outside of the fictional context, which it allegedly gains once it appears within the relevant context. Rather, what makes it fictional is the fact that, *regardless of its conventional meaning outside of the fictional context*, the proper assignment of meaning and truth-values to the piece of discourse is determined by the discourse itself once it appears in a fictional context. This interpretation assignment is given by means of a general cognitive

process, rather than a language specific, semantic process.

It is an independent claim of mine that empty names do not convey a proffered semantic (for more on the proffered /non-proffered distinction see Roberts 1998). This entails that utterances involving empty names are not intelligible *merely in virtue of* their semantics. But this should be no problem since my view takes them to be intelligible *mainly in virtue of* the general cognitive mechanisms and contexts that make use of them. I say "mainly" because I don't want to endorse the problematic claim that facts about the natural language being used play no role in this explanation. It seems plausible to think, for example, that certain facts about English do explain why 'Santa is coming tonight' is a meaningful piece of fictional discourse. What I claim is that such linguistic facts are not enough to deliver a satisfactory explanation. We need more to properly understand what goes on with the ordinary use of empty names. This extra element is, on my view, given by a general cognitive mechanism.

5.1.3   Pretense and imaginings

It is rather surprising that philosophical theories of empty names have simply ignored the psychological research on pretense. The latter is directly concerned with what speakers do, psychologically speaking, when engaged in fiction making. The data are evidently relevant. Yet, to my mind at least, not a single account in the philosophical literature on empty names has reflected upon this. This attitude is unfortunate not only because it is negligent but, most importantly, because it has precluded philosophical theories from benefitting in a rather substantial way. As I will show, a proper understanding of the cognitive mechanisms underlying human pretense is a key element of any adequate theory of empty names.

Generally speaking, there are two competing accounts of pretense available in the literature: behaviorism (see Lillard 1993a and 2001, and Nichols and Stich 2000 and 2003) and mentalism (see Leslie 1987, 1994, and 2000, and Friedman and Leslie 2007). Whether you are a mentalist or a behaviorist about pretense you are bound to posit some mechanism or other by means of which our cognitive apparatus "quarantines" the contents of pretense-attitudes from the rest. The idea, owed to Leslie 1987 and followed by Nichols and Stich 2000 (see p. 120 ss.) is a recognition of a rather simple fact: subjects need not believe what they pretend. When an adult pretends of a banana that it is a telephone she does so while still believing that it is not a telephone. Behaviorists like Nichols and Stich 2000 account for this by positing a separate "pretense box". Mentalists, specially Leslie 1987, do this in terms of a decoupling mechanism. Because it offers the most detailed account of the process, I make use of Leslie's decoupling account (more on section 5.2).

That's not all there is to the quarantine of pretend-attitudes. Pretended behavior usually

involves the use of props. When a child pretends of a banana that it is a telephone, that particular banana is central to her pretense. However, she doesn't pretend that *the banana is a telephone* but rather that *this is a telephone*. Similarly, when she pretends of an empty cup that it is full, she does not pretend that *the empty cup is full* but that *that cup is full*. According to Leslie 1987, the referential, truth-conditional, and existential commitments of the pretended representations are also quarantined. Thus, the goal of the decoupling mechanism (see Leslie 1987 and 1994) is to divorce the representation from its content.

It is this feature of pretense that I want to underscore. When describing the attitudes of speakers that use empty names, one must keep in mind that the mental states of these speakers are guided by mental representations and not by the states of affairs that these may or may not ordinarily represent. I believe this is a commonly ignored yet substantially relevant feature of the ordinary use of empty names. If the psychological research on pretense is correct, then what we need in order to explain the data is a proper understanding of the representational mental states involved. What is the right kind of attitude associated with the ordinary use of empty names?

5.2   The cognitive theory

Central theses

The cognitive theory of empty names is based on the following central theses:

Semantics: In virtue of semantics, all referential uses of names contribute their referent, and only their referent, to the content of utterances in which they appear.

Pragmatics: All referential uses contribute a non-proffered piece of information: that there is a referent of the name (some take this to be a presupposition).

Cognitivism: Referential uses of empty names are intelligible not merely in virtue of semantics and / or pragmatics, but mainly in virtue of the general cognitive resources they recruit (i.e., something like Leslie's 1987 "decoupling mechanism").

Representationism: Two object-directed attitudes, the content of which is cognitively determined (as above), may be directed toward the same object even if the associated representations have been (cognitively) assigned different referents.

I take the first thesis to be a familiar one, reminiscent of Millian views on proper names. I do not, however, have any particular interest in defending everything that goes under that

label (see Soames 2002 for a recent exposition). I do agree, however, that proper names are rigid designators and singular terms. But I also take these to be semantic features of names. My defense of this first thesis is mainly empirical: I take it to be the best explanation of the relevant psychological and neurological data (see García-Ramírez and Shatz forthcoming). I take the second thesis to be uncontroversial. There is, to my mind, a lot of agreement on this (see Heim 1982; Roberts, 1998 and 2004).

It is the third and fourth theses, **cognitivism** and **representationism**, that I take to be a contribution of the theory I want to propose. I will defend them in the sections to come.

The details

The cognitive theory of empty names comes in two parts. First, it assumes a referentialist semantics for names, according to which the only semantic contribution of a proper name is its referent. In doing so it allows us to extend the same semantic treatment to *all* referential uses of proper names, including empty ones. This is a great advantage and, to my mind, an important contribution of my theory to the debate on empty names. A consequence of this homogenous treatment is that assertions, belief reports, and fictional games that make use of empty names are *not* intelligible in virtue of their semantics alone. So we need the second part of the theory to explain how such uses *are* intelligible. Here I offer a psychological account of the mechanism involved in uses of empty names in a way that is independent of the semantics of the expressions used. Even though there is already a lot of empirical research on pretense, the account I offer is still preliminary. More needs to be said about the relation between pretense and the development of mental representational states. On this fluid view, when using empty names speakers assert, believe, report, and desire the contents of their imaginings, which are concerned with what I call "epistemically decoupled units" or EDUs for short. As I will hint at in section 5.4 the benefit of EDUs goes beyond the theory of empty names.

5.2.1 The decoupler

Leslie 1987 distinguishes among three kinds of mental representations: *primary* (e.g.,THAT THE CUP IS EMPTY), the function of which is to adequately represent the subject's environment according to the information received through perception; *decoupled*, which can be viewed as mental analogues of quoted sentences (e.g., THAT "THE CUP IS EMPTY"), the function of which is to divorce primary representations from their referential, truth, and existential conditions; and *metarepresentational*, which are higher order representations

(e.g., THAT SANDRA PRETENDS OF THAT CUP THAT IT IS EMPTY), the function of which is to offer proper objects of propositional attitudes and to adequately represent psychological states. It is important *not* to identify decoupled representations as the main product of the decoupler. The latter is a cognitive mechanism that is meant to deal with three different kinds of representations and its main goal is to yield a metarepresentation (on Leslie's view at least). Yet, it is part of the job of the mechanism to produce (as the result of a subprocess) decoupled representations.

Figure 5.1 (taken from Leslie 1987) gives a first approximation. Solid arrows correspond to input and output relations. Boxes correspond to independent mental representational states. The striped arrow corresponds to a mental function from *and* into mental states. This function is understood as the process of decoupling, which takes the primary representation produced on the basis of the initial "world" input and turns it into a metarepresentation.
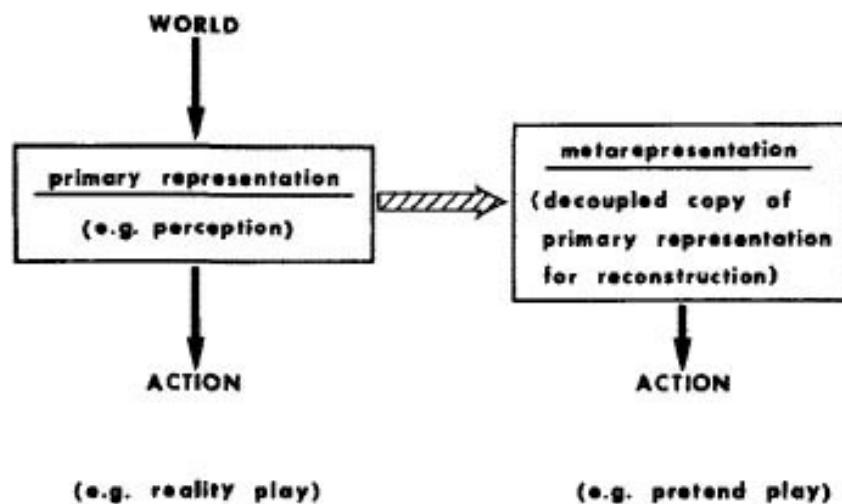

Figure 5.1 Primary and other representations

The decoupler is meant to be an independent mechanism, detached from general cognitive processes, and available for use in distinct tasks (e.g., mind reading). Its malfunctioning has been hypothesized to play a central role in autism, see Baron-Cohen, Leslie, and Frith 1985.

Figure 5.2 offers a more detailed description of the mechanism itself. As before, solid arrows correspond to input / output relations. This time, however, boxes correspond to distinct mental *processes*. There are five such processes: two main ones (i.e., perception and decoupling) and three subprocesses (i.e., expression raising, manipulating, and interpreting).

The mechanism of decoupling involves three distinct subprocesses. The first subprocess consists in *copying* the primary representation by divorcing it from its semantic values. One
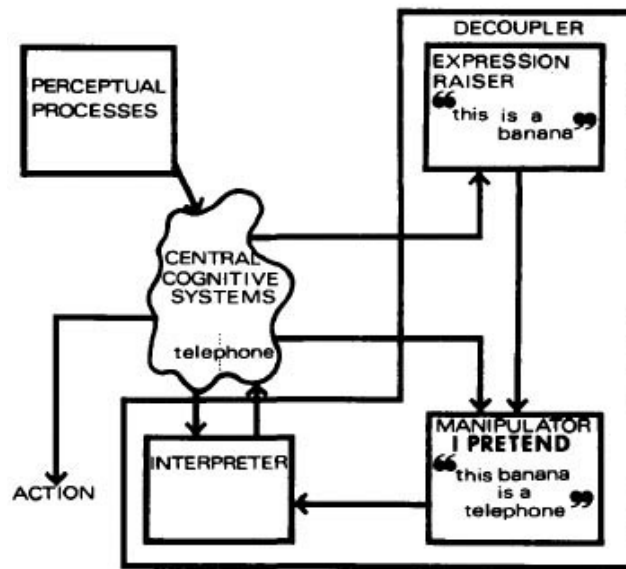
Figure 5.2   The decoupling mechanism

way to understand this is by analogy with the quote or mention of a given sentence, whereby both its meaning and truth-value are suspended.

The second subprocess consists of *manipulating* this decoupled representation by using information from central memory. This information may include, for example, rules of inference and general knowledge, as well as notions of attitudes. This relation with central memory is essential to understand how imaginings and pretense *develop* beyond the initial input. When children are told to pretend of an empty cup that it is full, they naturally behave as if its contents have been spilled when their peers turn the cup upside down. Perhaps of more general philosophical interest is the fact that this explains how decoupled representations can still figure in *inferential* processes. That is to say, the mental analogues of uninterpreted sentences can still play a role in the relevant inferential processes. The subjects may go from 'this cup is full' to 'there is water on this table' if they see that the relevant cup is turned upside down.

The third and last subprocess is that of *interpretation*, which consists in the assignment of the relevant meanings and truth-values in order to deliver a proper object of an attitude (e.g., believing, desiring, hoping, etc.) to central cognitive systems. As you can see in figure 5.3, the process of interpretation has access to central memory and, thus, to information gained through perceptual processes. In this way, the process may arbitrarily assign perceived objects as referents of the corresponding representation.
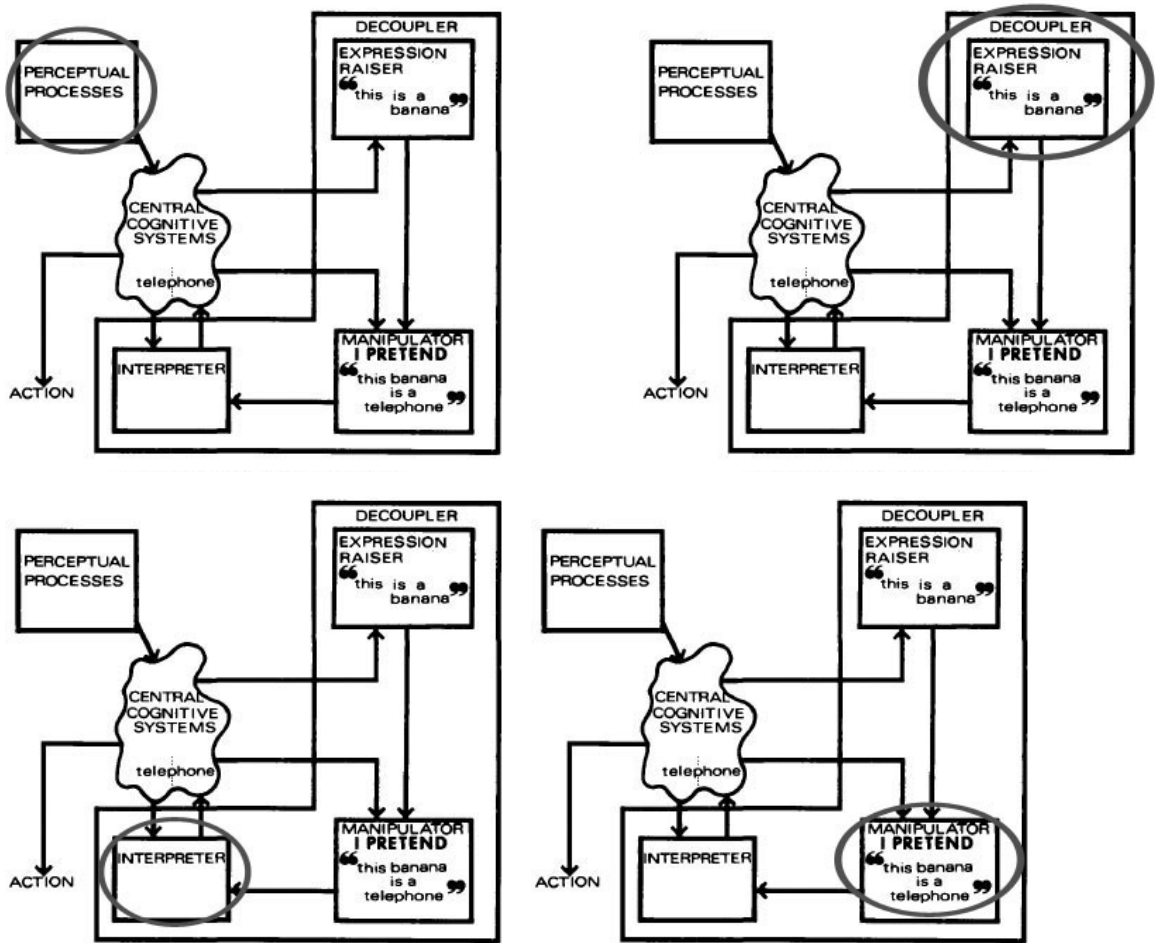
129

Figure 5.3   From upper left and clockwise: the three subprocesses of decoupling

5.2.2   Belief, pretense, and EDUs

According to Leslie 1987, the decoupler's main goal is to deliver objects of thought. But what exactly are these? Are they just traditional propositions? It is my belief that the decoupler must produce representational objects that are, in important respects, distinct from propositions. To see this we must take a closer look at "pretense" mental states and how they relate to "belief" mental states.

Humans are able to have mental representational states of two broadly different kinds: accurate and deviant. Beliefs, on the one hand, are of the former kind: they purport to represent the environment in an accurate and faithful manner (e.g., my belief that there's a chair in front of me). Pretense states, on the other hand, are of the latter kind: they purport to represent the environment in ways that the environment is not, and in some cases could not be (e.g. children pretend that broom sticks are horses or that bananas are telephones).

Yet, an important feature of human cognition is that there is no contamination of "deviant" upon "accurate" mental states. Humans normally do not believe what they pretend. Pretense states are taken to be quarantined by the decoupling mechanism above described.

There are important similarities among these two kinds of mental representational states. They both represent the environment. They both, also, guide behavior equally efficiently. To properly pretend that something is the case, one has to "play along" with the pretended representation of the environment (see Nichols and Stich 2002). Accurate and deviant representations are also similar in the way they form part of inferential processes. If one pretends of an empty cup that it is full, one will also pretend that its contents are spilled if the cup is turned upside down. Finally, belief and pretense (accurate and deviant representational mental states) also follow a similar developmental path (see Shatz 1994 on the development of language and ToMM from infancy to toddlerhood; see Bosco, Friedman, and Leslie 2006, on the development of pretense).

This underscores the differences between accurate and deviant representational mental states. Unlike belief, which is ordinarily caused by perception, pretense mental states are not a product of human perceptual processes. Pretense mental states have their origin in some or other representational assumption: e.g., that this broom stick is a horse. Belief and pretense also have different goals: pretense mental states are not in the business of offering faithful representations of the environment. And, finally, there is also a difference that can be described in terms of general uniformity: beliefs tend to be consistent with each other, pretense mental states can very well be explicitly inconsistent among each other.

Leslie 1987 has identified three different forms of pretense. "Imaginary object" pretense happens when subjects pretend that there is an object where there is none (e.g., that there is a teaspoon in front of you). "Fictional properties" pretense happens when subjects ascribe fictional properties to material objects they use as props (e.g., that a plastic doll has a dirty face). Lastly, "object substitution" pretense happens when subjects take an object to stand for another (e.g., when a broom stick is taken to be a horse, or when an actor is taken to be a character).

These forms of pretense require representational contents that seem to be at odds with propositions. Pretense mental states can be about what is impossible; furthermore, they can be about contradictions. To illustrate, consider the following passage from Borges' "The Book of Sand".

> He told me his book was called the Book of Sand because neither sand nor this book has a beginning or an end. He suggested I tried to find the first page.
>
> I took the cover in my left hand and opened the book, my thumb and forefinger almost touching. It was impossible: several pages always lay between the cover

and my hand. It was as though they grew from the very book.

"Now try to find the end".

I failed there as well. "This can't be" I stammered.

"It can't be, yet it is" the Bible peddler said. "The number of pages in this book is literally infinite. No page is the first page; no page is the last. I don't know why they are numbered in this arbitrary way, but perhaps it's to get one to understand that the terms of an infinite series can be numbered in any way whatever."

The passage is certainly comprehensible (and quite enjoyable). Yet, its understanding requires, at least, that one is able to pretend that some contradictions are the case. Whatever representational content the above passage conveys, it seems to be one that is *not* bound by the limits of logical space. In so far as propositions are bound by logical space, we need something else than just propositions to account for the pretended representational contents conveyed by Borges' work.

Consider now a theatrical representation. Peter and Paul observe two different performances of Shakespeare's "Hamlet" at the National Theater. Peter sees Daniel Day-Lewis performing as Hamlet, Paul gets to see Ian Charleson. As part of their understanding of the play, Peter and Paul are asked to pretend that ϕ: *that Hamlet is courageous*. It seems uncontroversial to say that both, Peter and Paul, share a common representational mental state directed towards ϕ. But things get complicated when we wonder what exactly ϕ is.

Both Peter and Paul are engaged in object substitution. Peter pretends that Day-Lewis is Hamlet, Paul pretends that Charleson is. According to Peter, whether or not Hamlet is courageous depends on whether Day-Lewis has a certain set of properties; according to Paul it depends on whether Charleson has them. So one and the same representational content, e.g., ϕ *that Hamlet is courageous*, has different truth-conditions or, if you prefer, different satisfiers depending on the theatrical performance. It seems, then, that ϕ is not a proposition, for propositions are defined by their truth-conditions.

We can find some common conditions of satisfaction. We can say that, independently of the performance, whether ϕ is the case depends on whether the relevant prop has the relevant properties. This would give us common truth (or satisfaction) conditions for Peter's and Paul's pretending that ϕ, but it still delivers the result that the representational content is *not* determined by the truth or satisfaction conditions. For it is clear that neither Peter nor Paul pretend that ψ: *that the relevant prop is courageous*; ϕ and ψ are not the same representation.

These rather brief remarks suggest that, whatever they are, the representational objects to which pretense mental states may be directed are importantly distinct from propositions.

132

Two instances of these representational objects may be considered to be of the same type, as with Peter's and Paul's pretense that ϕ, even though the instances have different referential (and hence truth) conditions. Peter's pretense takes 'Hamlet' to refer to Daniel Day-Lewis, Paul's to Ian Charleson. For reasons that should be clear in what follows, I will call these peculiar representational objects "epistemically decoupled units" or EDUs for short. In the following sections I offer a preliminary account of them by comparing them to propositions. Even though it is still in its initial stages, the account should be enough to show how it solves the problems posed by empty names.

5.2.3 EDUs and empty names

Of the three subprocesses mentioned in section 5.2.1 I want to underscore the third one. It is the goal of the interpretation subprocess to assign the reference, truth, and existential conditions of the corresponding representation. As a result of this process, empty names get assigned a relevant content. This relevance will be determined by the needs of the context. If, say, Sandra and Carla pretend of a doll that it is the Queen of England, then the referential term 'The Queen of England' in the representation 'The Queen of England has a dirty face' will refer to the doll, and the truth-value of the representation will depend on the fictitious properties that Sandra and Carla assign to the doll in question (e.g., having a face).

In all cases, by assigning an interpretation, the process delivers a meaningful representation that can now be the object of some attitude or other (e.g., the content of some belief or other). The resulting interpretation, however, is general cognitive, in that it requires the use of general purpose resources that go beyond purely linguistic ones. In this sense, the interpretation is *not* semantic. This illustrates the third central thesis of the cognitive theory:

Cognitivism: Referential uses of empty names are intelligible not merely in virtue of semantics and / or pragmatics, but mainly in virtue of the general cognitive resources they recruit.

The importance *qua referents* of the objects used as props in the corresponding games will vary from case to case. The game may include the rule that *this* and only *this* doll is to serve as a prop of the Queen of England. It may also be, as it seems to go with Santa these days, that many different objects may serve as a prop. In cases like this one, no particular object is required to be the referent of the relevant term. To illustrate, consider the following case.

Obama and McCain dress up like Santa and show up at Central Park, one on Monday, the other on Tuesday. Suppose further on that it's the first time Sandra and Carla meet with

Santa. Sandra goes on Monday, Carla on Tuesday. On Friday Sandra and Carla get together. They start wondering whether Santa will bring the gifts they expect. They do not wonder whether Obama or McCain will bring the gifts. The interpretation subprocess may very well have assigned Obama to be the referent of 'Santa' in the decoupled representation 'Santa will bring the gifts'. But it may very well have assigned any other object as well (e.g., a drawing, a plastic toy, or even a mental image).

The relevant object of the attitudes here is that which is common to both Sandra and Carla. They both entertain a full blown representation, with contents and truth-values assigned. But what is common to both is not the referent that each has assigned to their representations. Rather, what is common to both is the *decoupled representation* as it appears when divorced from its *referential* conditions. In cases like this, where the explanatory role is played by the (decoupled) representation and not by its assigned contents, I say that the attitudes are concerned with *epistemically decoupled units* (EDUs). This illustrates the fourth and last thesis of the theory:

Representationism: Two object-directed attitudes, the content of which is cognitively determined, may be directed toward the same object even if the associated representations have been (cognitively) assigned different referents.

To avoid confusion it will be important to tease apart two different kinds of contents that appear in the above thesis: (i) the object of the attitude, and (ii) the content of the representation. There is a sense in which they may both be called "content", one mental, the other truth-conditional. But we need to properly distinguish among these two "contents" if we are to make sense of object-directed attitudes—intentional or representational mental states—associated with empty names. It is important, then, to keep in mind that we are dealing with two distinct things: the object of the attitude (i.e., a representation) and the content of the representation (i.e., an interpretation). That said, what representationism claims is that two subjects may have the same attitude (e.g., belief) towards the same object (e.g., *Santa is coming tonight*) even if they assign different interpretations to that object (e.g., one assigns Obama to 'Santa', the other McCain). This is possible in virtue of the nature of the context in which the empty name is interpreted and insofar as the representation and interpretation in question are the result of something like the decoupling mechanism just described.

It is important to note, then, that on this view the cognitive interpretation *cannot* be simply generalized to apply to every single utterance. Whether or not the cognitive account applies depends on whether the situation demands that the subject make use of the relevant general cognitive resources. This, in turn, is always determined by the context (e.g., a

context of fiction making). These are the contexts where **representationism** applies. The cognitive theory of empty names claims that empty-name contexts generally make such cognitive demands.

To better understand the object-directed attitudes that I am talking about here, it will be useful to contrast them against good old "propositional" attitudes.

5.2.4 EDUs and propositions

I said the attitudes associated with empty names are concerned with a relevant EDU. For ease of explanation, I will call these "EDU-attitudes". What is, you may ask, the object toward which they are directed? It is a representation with referential, truth, and existential conditions assigned: a mental analogue of an interpreted sentence. In this sense, they are very similar to propositions. Perhaps one wishes to think of propositions as sets of possible worlds. Then one can also think of EDUs as sets of possible worlds: i.e., those stipulated by the interpretation assigned by the decoupling mechanism. Alternatively, one may wish to think of propositions as structured objects that mimic the syntax of the sentences used to express them, with a corresponding assignment of the referents of the terms used . Then one can also think of EDUs as structured objects in exactly the same way. The corresponding referents will, once more, be determined by the interpretation assigned by the decoupling mechanism. This naturally prompts the question: Isn't this just the same as a proposition? The short answer is no. The long answer requires some detail.

Differences

First and foremost, **representationism** is true of EDU-attitudes and, hence, it is also true of the objects to which they are directed: i.e., EDUs. According to this thesis, the referential conditions assigned to the EDU are not essential to them: a reference change need not result in a different EDU. If one wants to keep the notion of reference as a semantic one, then one can take the cognitive interpretation to assign "props" instead of referents. The example of Sandra and Carla's first meeting with Santa illustrates this very clearly. Sandra and Carla may associate different props with 'Santa' (e.g., Obama or McCain) and yet they may both have the same attitude (e.g., a belief) towards the same object (e.g., the representation *that Santa is coming tonight*) that results from the process of decoupling.

However, **representationism** is generally not true of the attitudes associated with non-empty names and, hence, it is generally not true of the objects they are directed towards.[4] If

---

[4]I use the hedge "generally not true" to be consistent with the central claim that whether or not a cognitive

135

these objects are propositions, then the thesis in question is generally not true of propositions.

Second, EDUs must be determined by something like the decoupling mechanism. This mechanism, in turn, is *cognitive-general*. This, however, is not true of propositions. Insofar as they are the objects of attitudes, propositions may be determined by task-specific linguistic mechanisms: e.g., by means of syntactic and semantic processing. Now, insofar as it is semantically determined, the referential conditions of a proposition are *essential* to it. The same is not true of EDUs. In virtue of this there are no two mental states that can be directed toward the same proposition and yet vary with respect to their referential conditions.

Third, this yields at least one important difference between propositional and EDU-attitudes: the interpretation of embedded clauses in mental state reports. According to the cognitive theory, the same sentence undergoes different *interpretations* depending on whether it denotes the content of an EDU or a propositional attitude—i.e., in the case of EDUs the referential conditions of the associated representation will be satisfied by props and may be open to referential substitution. What is important, though, is that this differential treatment is not explained by claiming that proper names may have different referents, but by pointing at the general cognitive mechanisms that produce the objects of such attitudes.[5]

Thus, propositions and EDUs differ in kind as representational objects. The former have their referential, existential, and truth conditions essentially. The latter do not. One way in which this difference appears is with respect to referential conditions. Propositions necessarily take referents. If the proposition is associated with a non-empty name, whatever unique object is the referent of that name it is so by necessity. EDUs may take props. If the EDU is associated with an empty name, whatever object the cognitive mechanism assigns as the value of the name, it is only as a prop and, hence, it is not so by necessity. Many other objects would do equally well, so long as they are assigned to the empty name by the relevant process of interpretation. The paradigm example is, once more, the Santa case with alternative props (e.g., Obama and McCain).

If we follow Rayo (forthcoming b) in distinguishing between formulas that are satisfied by possibilities and those that are satisfied by representations of possibilities, we can take propositions to always satisfy the former kind of formulas, while EDUs are able to satisfy the latter. Rayo (forthcoming b) makes this distinction by means of the dot notation ($\dot{x}$).

---

interpretation applies will depend on the context. Thus, even though I give no example here, the cognitive theory is open to the possibility of there being some use of a non-empty name that demands the use of the decoupling mechanism and that may be said to observe **representationism**. It is important to note, though, than in such case the name in question would not be used to convey its semantic content.

[5]This view on belief reports using empty names is wholly independent from the phenomenon of substitution failure. Empty name use and substitution failure are two clearly different phenomena. Elsewhere (see García-Ramírez, forthcoming a) I defend a predicate-transfer view of substitution failure that can be easily applied to the case of belief reports that exhibit both phenomena: substitution-failure while using empty names.

Thus, on this view, nothing can be said to satisfy the undotted sentence below (since there is no object that is identical with Santa), but there is an EDU that can be said to satisfy the second, dotted sentence (e.g., one that assigns Obama as a prop for Santa).

$$\exists(x)(Santa = x) \wedge C(x)$$
$$\exists(x)(Santa = \dot{x}) \wedge C(\dot{x})$$

The difference is simple: ($\dot{x}$) takes us from possible objects to representations of these possible objects. There are no possible objects that are Santa. But there are lots of them that can represent Santa. It goes without saying that Rayo's account is importantly different from the cognitive theory I defend. His solution to the problem of empty names consists in taking semantics (in general) to quantify over representations of possibilities instead of possibilities. His account is, thus, semantic, not cognitive.

Similarities

I hope to have said enough about the differences between EDU-attitudes and propositional attitudes. In order to fully characterize the former it is important to also note some similarities.

There is, first, an important psychological similarity. Even though they differ in virtue of being directed toward different kinds of objects, EDU and propositional attitudes are both guided by objects they are directed toward. If I believe that Santa is coming—i.e., if my belief state is directed to the relevant EDU— and I believe that Santa loves cookies, I will go and leave cookies by the tree. If I believe that Obama is coming—i.e., if my belief state is directed to the relevant proposition— and believe that Obama likes cookies, I will go and leave some cookies by the counter. Thus, EDU and propositional attitudes cause behavior in exactly the same way.

Second, there is an important epistemological similarity. Both EDU and propositional attitudes may be said to be adequate depending on the actual world. If, for example, the relevant attitude has a mind-to-world direction of fit, it will be adequate (or true) depending on what goes on in the actual world and regardless of whether its content is a proposition or an EDU (more on this later). This seems to be the case with belief states. Whether they are directed to an EDU or a proposition may be irrelevant when it comes to judging their epistemological adequacy. If I truly believe (as opposed to merely pretend) that Santa is coming tonight, my belief will be said to be epistemically inadequate even if it does have an object (i.e., an EDU) as its content.

As for attitudes that are clearly concerned with fictions, such as pretense, their conditions of adequacy will vary depending on the context. Whether and how the adequacy of these

attitudes depends on the actual world is determined by the fictional context itself. For example, if we pretend of a particular doll that it is the Queen, pretending that the Queen has a dirty face may be adequate depending on the properties the doll actually has.

What we have, then, is a two-fold distinction between kinds of mental states: in terms of direction of fit and in terms of kind of representational content (see Table 5.1).

Table 5.1   Kinds of Mental States

Mental States	Direction of Fit	Referential Commitment	Attitude
Beliefs, hypotheses, etc.	Mind-to-World	Fixed	Propositional
Beliefs, hypotheses, etc.	Mind-to-World	Flexible	EDU
Desires, intentions, etc.	World-to-Mind	Fixed	Propositional
Desires, intentions, etc.	World-to-Mind	Flexible	EDU

Individuation

There are two defining properties of EDUs: (i) their referential conditions are not held fixed (i.e., their referential terms may substitute reference without changing the representation); and (ii) they are the product of the decoupling mechanism.

Given the truth of **representationism**, EDUs do not have their referential conditions essentially—they exhibit what Leslie 1987 calls "object substitution". Two subjects may have attitudes directed toward tokens of the same EDU even if they assign different props to the same corresponding referential term. Sandra and Carla, for example, both believe that Santa is coming tonight. They both have a belief state directed towards the same object: i.e., a token of the same EDU.

But EDUs are *also* essentially the product of the decoupling mechanism described in section 5.2.1. As mental representations, the relevant tokens and types of EDUs should be individuated as products of the relevant cognitive processes. (For a similar view that individuates representational content according to the cognitive mechanism responsible for its production see Leslie 2008.)[6]

Through the first subprocess of copying, the "expression raiser" (see page 128) will deliver the same uninterpreted sentence (e.g., 'Santa is coming tonight') in both cases. Hence, for two objects of thought to be tokens of the same EDU they must correspond to the same uninterpreted sentence in the language of thought. Through the second subprocess the "manipulator" assigns inferential relations to the copied representation. Hence, for two objects to count as tokens of the same EDU they must have, at least, equivalent inferential

---

[6]On Leslie's 2008 view, generics (i.e., representations of the form 'K's are F's') are essentially the product a default mechanism of generalization.

patterns (this may be determined by the context). Finally, through the third subprocess the "interpreter" assigns reference, existence and truth conditions by assigning referents and meanings to the representation delivered by the previous subprocesses. We know the reference conditions are not essential. But the same is not true of the existence and truth conditions. EDUs should also be individuated in terms of these.

We have, then, four criteria for EDU individuation: reference substitution, correspondence to uninterpreted sentences, inferential pattern, and existential and truth conditions. All of these are determined by the decoupling mechanism.

I think this machinery is capable of solving all the relevant problems having to do with ordinary use of empty names. I will present the solutions in the following sections. Let me conclude the present section with two disclaimers. First, the cognitive account I have described is meant to work on a case by case basis. It cannot be systematically generalized to apply to any use of any term whatever. This precludes the account from overgenerating. Whether the cognitive mechanism is in use and how it assigns values to terms is mostly up to the context. Second, the distinction between EDUs and propositions (and between their corresponding attitudes) is theoretical. I do not take speakers to be aware of such distinction, or to know that they ascribe different kinds of attitudes depending on whether they use empty names or not.[7]

5.3   Problem solving

5.3.1   Homogeneous semantics

The label "empty names" comprises three distinguishable kinds: fictional / mythical names (e.g., 'Santa', 'Hamlet', 'Pegasus'), mistaken scientific hypotheses (e.g., 'Vulcan'), and (perhaps) hallucinations—though this might be more about empty demonstratives than names. The cognitive theory gives them all the same semantic treatment: they all lack a proffered semantic value. Yet, they all convey the presupposition that there is a unique object to which they refer (see page 126). A proper understanding of their ordinary use requires that they be assigned an interpretation by means of something like the decoupling mechanism.

---

[7]The cognitive theory of empty names does not require that speakers know there is a distinction between EDU and propositional attitudes, just like Leslie's 1987 meta-representational account of pretense does not require that subjects know there is a distinction between representations and meta-representations.

5.3.2 Meaningfulness

This helps us solve Reimer's 2001a "intuition problem", according to which speakers may think that the sentences they use are meaningful even though they use empty names. The case of speaker and hearer is easy. They take the speech acts to be meaningful in virtue of cognitively interpreting them to be so. To properly engage in referential uses of empty names, according to the cognitive theory, speakers must make use of extra-linguistic, general cognitive resources such as the decoupling mechanism. Speakers (and hearers) find referential uses of empty names to be meaningful because they imagine these names to have a referent. This also explains why speakers (and hearers) may find such utterances to be truth-evaluable.

It may be, however, that third parties also find such speech acts meaningful. Someone who is not aware of Shakespeare's work may nevertheless find an assertion of 'Hamlet is courageous' to be meaningful without knowing that 'Hamlet' is a fictional name. She doesn't know what the conversants are talking about, yet she finds the utterance meaningful. Perhaps this third party cannot be said to make use of the decoupling mechanism. Perhaps she finds the use of 'Hamlet' meaningful without having to exercise her imagination to produce an interpretation. This is possible, I think, only if we take the third party to interpret the utterance of 'Hamlet is courageous' as meaning something like "There is a unique thing that is courageous" or like "The thing (person) the conversants are talking about is courageous". This is not explained by appealing to the decoupling mechanism, but only because it is already explained by appealing to the presuppositions conveyed by the use of 'Hamlet': i.e., that there is a single object to which uses of 'Hamlet' refer. Third parties may find uses of empty names meaningful, without having to come up with an associated referent, in virtue of the presuppositions conveyed by all referential uses of names. All this is part of the cognitive theory of names (see section 5.2, page 126).

5.3.3 Fictional discourse

The account is also *compatible* with the distinction between fictional and non-fictional discourse defended in section 5.1.2. On this view, fictional uses of empty names are self-granting in virtue of prescribing their own interpretation—the fictional context itself determines how the relevant speech act is to be assigned reference, existence, and truth conditions. Since this is a cognitive, and not a linguistic, distinction it can be suitably generalized to account for the fiction / non-fiction distinction. So long as we take objects of fiction to be representations (and I see no reason against doing so), they can be said to

prescribe their own interpretations. On this count the cognitive theory fares better than Sainsbury's 2005 and 2010 RWR theory (see section 5.1.2).

5.3.4 Psychological adequacy

The cognitive theory is consistent with the available evidence on pretense and fictional reasoning. It partly stems from such research. But, does it face similar empirical problems as Sainsbury's RWR theory? (See section 5.1.2.) According to the cognitive theory, empty and non-empty names are *semantically* on a par: all they require is an understanding of reference. There are reasons to think this understanding is prelinguistic (see García-Ramírez and Shatz forthcoming). As such, the theory is compatible with the evidence suggesting that proper names are understood as early as 6 months of age and, hence, a long time before there is an understanding of identity statements.

The theory does claim that empty names, in particular, demand more than just linguistic resources. There are independent reasons to think that competent use of these resources (i.e., something like the decoupling mechanism) doesn't appear until some time between the second and third year of age (see Friedman and Leslie 2007). The cognitive theory predicts, then, that competent use of empty names will not appear until some time between the second and third year of age, once the decoupling mechanism is in use. There is, to my mind, no evidence against this prediction, but there is some evidence that speaks on its behalf. On this view, competent use of empty names appears at about the same age that children begin understanding and engaging in acts of pretense, and it is also about the same age when they understand identity claims and false belief. All these notions seem important to understand what speakers do with empty names.

5.3.5 Non-fictional empty names

I have been speaking mostly of *fictional* empty names, but the cognitive theory is good enough to account for referential uses of both fictional and non-fictional empty names. On this view, there is no difference in kind between fictional (pretense-like) uses of empty names and, say, serious (scientific-like) uses of them. There are, of course, important *contextual* differences between Le Verrier's hypothesizing *that Vulcan is a planet that is responsible for changes in the orbit of Mercury*, and an adult's pretending *that Santa is coming tonight*. Yet, according to the cognitive theory, there is no linguistic difference and they both require the use of the same general cognitive mechanism.

The important difference between fictional and serious uses of empty names seems to be, then, attitudinal. In both cases, the theory takes the speaker to have a mental state directed toward an EDU, but the way they are directed differs. When used fictionally, empty names are associated with a fictional or pretense-like attitude. When used seriously, empty names are associated with a more epistemically committed kind of attitude (e.g., a belief). As such, empty names and hence, EDUs can be associated with mental representational states with a mind-to-world direction of fit.

Consider, for example, 'Vulcan'. Le Verrier postulated the existence of a planet in order to explain the disturbances in the orbit of Mercury. He called this hypothetical planet 'Vulcan'. It turned out that there was no such planet and so 'Vulcan causes disturbances in the orbit of Mercury' turned out to have no truth-value determined solely by semantics. We can, I believe, reconstrue Le Verrier's thoughts as imaginings and describe his intentions accordingly. It could go as follows: we can conceive that Vulcan causes disturbances in the orbit of Mercury; to believe *this* simplifies our theory and hence, is acceptable until we find evidence to the contrary. Le Verrier did believe the representation he imagined (i.e., an EDU), so he had an attitude with a mind-to-world direction of fit. Further evidence showed Le Verrier's hypothesis to be mistaken. No matter how well he imagined it, his belief turned out to be false.

It is important to note that not all *beliefs* directed towards EDUs are automatically unwarranted. Whether they are depends on the world and hence, on the evidence. Le Verrier also hypothesized that Neptune caused disturbances in the orbit of Uranus. He had no more evidence of the existence of Neptune than he had of Vulcan's; as with the latter, he was imagining an EDU and accepting it (or believing it) for theoretical purposes. It turned out that he was correct in believing such a representation.

Presumably what happens once the subject finds evidence is that the relevant representation (i.e., initially an EDU) gets assigned an actual referent. At this time the information gained through perceptual means is processed by the central system (see figure 5.2). Now, once the relevant term, say, 'Neptune', gets its reference fixed with the information gained through perception, the resulting representation lacks both of the defining features of EDUs: it is no longer a product of the decoupling mechanism and its referent is not flexible or substitutable; it is what Leslie calls a "primary representation" (see page 127) or what I have here called a "proposition" (see page 135).

This proposition shares an important causal history with its EDU ancestor, and they share some of the individuating elements for representations: i.e., they correspond to the same uninterpreted sentence in the language of thought, and have equivalent inferential patterns and perhaps also existence and truth conditions. This allows us to say that Le Verrier kept

having the "same" belief throughout.

5.3.6 Errors and hallucinations

The case of errors is easily explained as well. Consider the following example:

> Andy is lost in the desert; he is dehydrated. After hours of looking for someone, he hallucinates a dark spot by a lake, at a distance. He says to himself, "He is by the lake". The sentence 'He is by the lake' has no semantically assigned referent, yet Andy does have a mental state with content: because of his hallucination he comes to believe *that he is by the lake*.

What's the value assigned to 'he' in that representation? I think there's one good candidate: the dark spot that Andy hallucinates. Does that spot exist? I think it does: it's something in Andy's mind. Remember, anything may be assigned as a referent by the subprocess of interpretation. What doesn't exist is a person to whom Andy falsely believes he is referring. In this particular case, Andy's hallucinatory state turns out to be a false belief (for some independent reasons to understand hallucinations in terms of false belief see Byrne 2009).

5.3.7 Belief reports

Appealing to EDU attitudes also accounts for the problems associated with mental content (see section 5.1.1). EDU attitudes are directed towards complete, full-blown, objects of thought: an EDU with reference, existence, and truth conditions. As a result, they are fine-grained enough to be psychologically explanatory, qua content, and they are as close as anything can be to a singular proposition without being such: they are representations that, if they were not detachable from their referential conditions, would represent singular propositions. But how exactly does this work for belief reports? Here are some examples:[8]

(1a) Sally believes that Santa is coming tonight.
(1b) Le Verrier believed that Vulcan was a planet.
(1c) I used to believe that Santa exists, but now I don't.
(1d) She used to believe that Santa exists, but now she doesn't.

These are all "serious" uses of empty names. Fortunately, the fact that they are serious does not constitute a problem for my account. The cognitive theory is not a pretense theory. It does not take speakers to be always pretending (or to always ascribe pretense states) when

---

[8]I am not considering belief reports that exhibit substitution failure. See footnote 5.

they use empty names. This, I think, is a great advantage of the account. As a result, it can assign interpretations regardless of whether speakers pretend or not. What the theory claims is that assertions using empty names either assert or ascribe an EDU attitude with an EDU as content. So the examples in (1) all ascribe EDU attitudes towards EDUs, according to my account. These objects of thought are full-blown mental contents with reference, existence, and truth conditions assigned. So there's nothing problematic with them. (1a) and (1b) ascribe the EDUs in (2a) and (2b) to Sally and Le Verrier correspondingly, while (1c) and (1d) describe how a given subject changed her mind from having a belief like EDU-attitude towards the EDU in (2c) to lacking it. If we want, we may use an analogue of Rayo's dot notation $\dot{x}$ to signal that a given content of a mental state is an EDU.

(2a) *that Sȧnta is coming tonight*
(2b) *that Vuḷcan is a planet*
(2c) *that Sȧnta exists*

It should be clear that these are unproblematic cases.[9]

5.4   Advantages

I like to see my theory as having a mix of strength and parsimony. As I have shown, it illuminates the many problems associated with the ordinary use of empty names, and it does so in a unique way. Unlike all other theories, this one is compatible with a satisfactory account of the fiction / non-fiction divide as well as with the evidence from language and cognitive development.

So much for the strength. A very attractive part of the theory, to my mind, is its simplicity and frugality. It offers a homogenous semantics for all referential uses of names, and it does not require us to believe in the existence of fictional or any other special entities. It does appeal to something like Leslie's 1987 decoupling mechanism and the kinds of mental representations associated with it. But this resource is one that we can have for free: we will need it anyway if we want to account for the psychological process of pretense, fiction making, and reasoning about false belief.

The cognitive account is also useful for purposes that go beyond the case of empty names. Here are some examples.

---

[9]If you are worried about the existential claims in (1c)-(1d) see section 5.5 on negative and positive existential claims.

5.4.1 Metaphysics

There is a sense in which, when using a name like 'Santa' or 'Hamlet', speakers are said to be talking about *the same thing*. This apparently happens even without pretense, for consider the Shakespeare scholar whose life is dedicated to the study of *Hamlet*. What is she studying? Works of fiction, of course. But, what are these? Fictional entities? Some, like Salmon 1998, are happy to say: Yes, these are fictional entities. The theory I have proposed suggests, however, a negative answer: No, fictional works are not fictional entities, nor are they about fictional entities.

Fictional works are representations. But there is no need to go from this to accepting the existence of what they pretend to represent. What matters, if the theory is correct, is not what they represent but the works themselves *qua* representations. So Shakespeare scholars study representations and this is exactly the kind of thing that they talk about when they are talking about *Hamlet*. The same happens, the theory suggests, with all other fictional representations including fictional names. There is nothing behind 'Santa', 'Pegasus', or 'Venus'; what speakers talk about when they use these names, that *very same thing* they share information about, is the representation itself, the contents of their own imaginings or, as I put it, some EDU or other. (I have already explained how two subjects may be said to have mental states directed toward the same EDU; see page 138.)

The cognitive theory explains how speakers manage to be concerned with the same representation (or mental representation) in terms of the content of the mental states. Another way to do this is to focus on the causal-historical chains of linguistic practice. If scholar A's use of 'Hamlet' is causally related to scholar B's use in the relevant way—they are both about the same representation, tokens of which they have seen, read, or heard—then scholar A and scholar B are talking about the same thing.[10] So, I propose, following Walton 1990, that we stop talking about fictional objects and start talking about representations (some of which are mental). To get an idea on mental representations, see Stich and Warfield 1994, and for representations in general, look at Walton 1990.

5.4.2 Problematic assertions

It has been objected against pretense accounts of empty names that they cannot explain serious uses of empty names. To illustrate, consider (3a)-(3d) (owed to Sainsbury 2010 and Eric Swanson in conversation):

---

[10]For a similar causal account of "similar content" see Everett's notion of THIN-ABOUTNESS in Everett 2000. His general theory of empty names, however, is one I disavow.

(3a) The Greeks worshipped Zeus.

(3b) Holmes is famous.

(3c) Anna Karenina is smarter than Emma Bovary.

(3d) Santa is a fictional character.

Since it is not a pretense theory, the cognitive theory has no problems dealing with these examples. It claims that serious assertions of (3a)-(3d) convey EDUs, which may very well be accepted or believed by whoever utters them. As I said before, the fact that they are EDU attitudes does not make them less (mind-to-world) committed. Suppose for the moment that worshipping is a matter of believing. Then on my view, someone who utters (3a) may very well believe a higher order EDU that itself attributes a belief-like EDU attitude (or a set of these) to ancient Greeks. As such, the attitude ascribed by uttering (3a) may or may not be warranted depending on the evidence we have about the attitudes of ancient Greeks. Similar treatments are available for assertions of (3b)-(3d).

5.4.3 Explaining intentional inexistence

According to Brentano 1874:

> Every mental phenomenon is characterized by ... the intentional (or mental) inexistence of an object, and ... reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something as object within itself. [Brentano, 1874, p. 88-89]

This paragraph has been highly influential in the philosophical tradition. It presents two theses that almost everyone agrees with, yet there seems to be no agreement on how to cash them out (for more details see Jacob 2003). The first thesis defines mental states as intentional states: all and only mental states are *about* something, they *tend* to something distinct from themselves. The second thesis takes mental states to be able to be about things that do not exist (i.e., intentional inexistence). This second thesis seems to be the most problematic: it suggests that we need to believe in so called "intentional objects", things which mental states are about and yet do not exist. Some feel uncomfortable postulating things that are but do not exist. If possible, we should avoid making such assumptions.

The cognitive theory explains how speakers can be said to have thoughts about Santa, Hamlet, Vulcan, etc., by having thoughts about some or other EDU. Thus, Brentano's second thesis is satisfied: by having mental states directed toward objects that are produced by the decoupling mechanism, a speaker may be said to have thoughts about something that does not exist. There is no need to look for special entities that do not exist. All we need is a

relevant interpreted representation the interpretation of which does not require anything but existing objects: i.e., an EDU. These objects exist; they are mental representations. Thus, Brentano's second thesis is verified without endorsing any extra ontology. Furthermore, EDUs are distinct from the mental states that are directed toward them. Thus, Brentano's first thesis is satisfied. The cognitive theory explains how we can have thoughts about things that do not exist without there being non-existing things and while distinguishing the content of the thought (i.e., an EDU) from the thought itself (e.g., a belief directed toward an EDU).

I think this account of intentional inexistence does better than its rivals. But I don't have the space to defend it here. For some independent evidence on behalf of this account see Gómez 2008, who offers an evolutionary account of pretense by associating it with the appearance intentional inexistence. For an alternative account see Kriegel 2007.

5.4.4 Metanormative theory

The cognitive theory explains how sentences using vacuous terms can be meaningfully used. It is also explains how subjects can have mental states directed toward full blown contents that may be conveyed by uses of those sentences and hence, how they can be guided by entertaining the corresponding objects of thought. As such, it is a general theory, not only a theory of empty *names*. It might prove to be useful when trying to understand the content of statements and mental states associated with expressions which we may have independent reasons to think are vacuous (perhaps because we have no evidence of the existence of their referents or no need to postulate them).

Contemporary debates in metaethics have raised the question as to whether moral and (perhaps) aesthetic terms are vacuous or not—and so, as to whether sentences such as 'Killing babies is morally wrong' have a semantically determined content. If we do have independent reasons to think that such is the case (i.e., that moral and aesthetic terms fail to have an actual referent, content, or truth condition), the cognitive theory may prove to be useful in explaining why those sentences may be used in meaningful ways, how subjects may have mental states with full blown proposition-like contents associated with such statements, and how ascribing mental states with such contents may help understand the behavior of these subjects.

147

5.5 Negative existentials

There are some uses of what appear to be empty names that I have not dealt with. These are uses that seem to be referential, but are such that we are not justified in associating any object with the name (not even as a prop). One is not justified, in these cases, in assigning EDUs to the relevant speech acts. These are uses of empty names in negative existential constructions of the form 'X doesn't exist' where 'X' is an empty name. The dialogue in (4a) gives a good example of this kind of use:

(4a) L: Lock your doors, Santa's coming!
E: This is the stupidest thing I've seen in a while, but what would I expect from Spam Land?
M: How can a kid be afraid of the big red guy?
N: This article is complete BS. How could anyone be afraid of Santa? Santa doesn't even exist.
(found online)

I think the cognitive theory can accommodate these cases. There are independent reasons to think that negative existential assertions *do not* involve referential uses of empty names. On my view, we are better off understanding negative existentials as involving *self-referential* uses, i.e., where the name, say, 'Santa', is used to refer to itself. Thus, on this view, the speaker conveys some metalinguistic information about the name used in the negative existential sentence. Strictly speaking, this use of an empty name is not empty: it refers to itself. Yet it is because ordinary *referential* uses of the same name are empty that negative existential assertions are true. I will explain how this works in the remainder of the paper.

5.5.1 Positive and negative uses

Let me begin with some useful taxonomy. There are, generally speaking, two kinds of uses of empty names within negative existential constructions. On the one hand, there are cases where the speaker *does* intend to use the name in a referential way. These are cases where the speaker intends to convey, pragmatically or otherwise, the non-proffered information associated with all names (see the theses on page 126): i.e., that there is a referent of the name. I dub such uses of negative existential assertions "positive uses". Here are some examples (found online) of such uses:

(5a) Check node type doesn't already exist when installing a module.
(5b) Free speech doesn't always exist online.

(5c) The Higgs boson particle doesn't already exist.

(5d) There are many things that don't exist: Pegasus, Santa, Hamlet.

(5e) According to the authoress, diseases can be grouped into two categories: those which come from the Gage world, and those internal to the group... Yet, this distinction *doesn't sometimes exist* and they turn to any possible cure in order to recover from the disease.

Positive uses are not problematic for the cognitive theory. Since the speaker is presupposing that there is a referent, the hearer will be justified in interpreting her utterance by assigning a relevant object as the referent of the term. If the name in question is empty, the cognitive theory explains how the hearer can come up with a useful interpretation by means of something like the decoupling mechanism. There is nothing odd in interpreting positive uses, such as those in (5) above, by taking the relevant name (or, in general, an NP) to have a referent. What requires some non-ordinary interpretation is the speaker's use of the English expression 'exists'.[11] Consider, for example, (5d). It seems clear that the speaker means to say that there are some things that are mythical or fictional. Another important feature of the positive uses is that they admit Positive Polarity Items (PPIs) (e.g., sometimes, not always, etc.) as in (5a), (5b), and (5c).

On the other hand, there are cases of negative existential assertions where the speaker does *not* intend to convey the information that there is a referent associated with her use of the relevant name. In these cases, the speaker cannot be said to presuppose that the name she uses has a referent. I dub these "negative uses" of negative existential assertions. An example of such a use appears at the end of the dialogue in (4a). Given the speaker's intentions, the hearer does *not* seem to be justified in interpreting her utterance by assigning any referent to the name. It would be odd to interpret N's assertion in (4a)—i.e., *Santa doesn't even exist*—as referring to Santa and claiming of it that it doesn't even exist. The speaker seems to be rejecting the very idea that the relevant NP (e.g., 'Santa') has a referent. An important feature of negative uses, like that in (4a), is that they seem to admit only Negative Polarity Items (NPIs) (e.g., even, never).

Negative uses of negative existentials are not readily explained by the cognitive theory. In so far as it seems incorrect to take the speaker to be referring to some object or other it seems incorrect to take the cognitive theory (as I described it) to apply to this case. It is negative uses, then, that are left to be explained.

To reach a proper explanation, we first need to be clear about the goals that speakers have while using negative existentials in this manner. The speaker in (4a) can't be properly

---

[11]In terms of conventional meaning the English expressions 'exists' and 'there is' are equivalent. Yet, there are several ordinary uses where they are not. The sentence "There is nothing you can do about it" is clearly not equivalent to "What you can do about it doesn't exist". For more on the ambiguity of the English expressions 'there is' and 'exists' see Fleming and Wolterstorff 1960.

said to be presupposing that the relevant NP has a referent. Rather, it seems that what she wants to do is close off all talk about Santa. Here are some other examples that support this view (all found online).

(4b) Web 2.0? It doesn't exist!

> My problem is not with the characterization of the components of Web 2.0. It is the implication inherent in the very livery, "Web 2.0," that I just dont get. How did we get here? First of all, Web 2.0 is a marketing slogan. The problem I have with this "Web 2.0" slogan is that it is a contrivance, meant to imply a unified movement or wave toward a better Web (...) That is, until some clever marketers wanting to charge a fortune for you to attend their conferences dreamed this one up. Dreamed Web 2.0 up as a nice-off-the-tongue, easily memorable descriptor for come to our conferences, learn about whats hot. Or as Wikipedia puts it, "Skeptics argue that the term is essentially meaningless, or that it means whatever its proponents decide that they want it to mean ..."

(4c) I say dark matter doesn't exist.

> In a study detailed in the Nov. 21 issue of the *Monthly Notices of the Royal Astronomical Society*, [John Moffat and Joel Brownstein say] their Modified Gravity (MOG) theory can explain the Bullet Cluster observation. MOG differs from other modified gravity theories in its details, but is similar in that it predicts that the force of gravity changes with distance. "MOG gravity is stronger if you go out from the center of the galaxy than it is in Newtonian gravity," Moffat explained. "The stronger gravity mimics what dark matter does. With dark matter, you take Einstein and Newtonian gravity and you shovel in more dark matter. If there's more matter, you get more gravity. Whereas for me, I say dark matter doesn't exist. It's the gravity that's changed."

It seems that in (4b)-(4c) the speaker / author is negatively using the relevant referential expression—she can be appropriately understood to accept that the NP in question does not have a referent. It also seems that an adequate account of the goals of these assertions includes the expectation that there be a relevant shift in the conversation. The speaker may want to shift the conversation so that the relevant NP is no longer used or that the talk about the alleged denotation of the relevant NP be closed off. More specifically, with (4b) we are asked to stop talking about Web 2.0; and with (4c) certain astronomers ask that their peers stop talking about dark matter and focus more on gravity. Since these are the ones left to be explained, from now on I will use "negative existentials" to refer only to negative uses of the existential construction.

With this in mind, I want to defend the following hypothetical claim: if we assume that the relevant NP used in a negative existential is used self-referentially, we can avail ourselves

a metalinguistic account of their use. To use a name, say 'Santa', self-referentially is to use it not to refer to Santa, but to refer to the name 'Santa'. Though semantically obvious, the distinction is *not* obvious usage-wise. As Geurts 1998 puts it:

> All things considered, language users don't seriously try to make formal distinctions between linguistic expressions and their names. Instead they simply use an expression to refer to itself, whenever the need arises. It follows from this that ALL expressions of any language are equivocal in a way: besides their ordinary meaning(s), they can also be used to designate themselves. [Geurts, 1998, p. 291.]

Let me now describe the metalinguistic interpretation I have in mind.

5.5.2 Metalinguistic interpretation

There are, to my mind, two different ways in which a negative existential may be used to achieve its conversation-shifting goal. The most straightforward way to do it is by uttering a negative existential that makes use of the relevant NP. On this view, a speaker can convey the information expressed by (6a) by uttering (4a) in the relevant context.

(4a) Santa doesn't even exist.
(6a) 'Santa' doesn't even name something.

But one can also achieve the same goal by means of anaphoric reference as E does in (6b).

(6b) L: Nessie is carnivorous.
    E: No, she isn't. She doesn't even exist.

On my view, these two uses of negative existentials may benefit from two different mechanisms that allow the audience to recover the metalinguistic interpretation from the conventional semantics of the sentence used. These are the mechanisms of *meaning transfer* and *presupposition accommodation*. Briefly put, negative uses such as that in (6a) are taken to involve selfreferential uses of their NPs, while those like (6b) are taken to deny a metalinguistic presupposition of the utterance.

This metalinguistic view is not new, it has its ancestors in Stalnaker 1978, Walton 2000, and apparently also Clapp 2008. Even though I agree with the spirit of these proposals, I believe they have an inconvenience: none of them tells us how it is that hearers may recover this non-literal (metalinguistic) interpretation from the literal semantics of the uttered sentence. My main contribution to this tradition will be to point at the mechanisms by means of which speakers may let hearers recover such interpretation.

5.5.3 The mechanisms

Meaning transfer

Meaning-transfer was initially proposed and developed by Nunberg 1979 and 2004.[12] The mechanism is meant to explain how speakers may use expressions to convey information that differs from their conventional content. For example, speakers may use (7a) to convey something more like (7b).

(7a) I am parked out back.
(7b) I am the owner of a car that happens to be parked out back.

Nunberg 1979 and 2004 describes two conditions on predicate meaning-transfer: **functionality** and **noteworthiness**. The former requires that the context include a salient function mapping the predicate used onto the relevant (transferred) property by means of a function mapping the ordinarily associated property onto the new one. The latter requires that the transferred property be useful to classify its bearer for the purposes of the context. It seems to me that every context in which a proper name is used includes a salient function mapping the predicates associated with the referent of such name onto the property of naming the object that has such property. For example, in a context where speakers say things like 'Santa is red', there will be a salient function mapping the predicate 'being red', which is associated with the referent of 'Santa', onto the property of NAMING SOMETHING RED, which is associated with the name 'Santa'.

In the case of negative existentials, there will be a function mapping 'exists', which may be said to be associated with the property of BEING SOMETHING, onto the property of NAMING SOMETHING. It seems also that in such contexts it will be useful to classify the relevant NPs (e.g., 'Santa') depending on whether they do or do not have a referent. Thus, it will be useful to classify 'Santa' depending on whether it does or does not name something. Thus, the property of NAMING SOMETHING will be a noteworthy one in the context where a negative existential may be felicitously uttered. With these ingredients in hand, the mechanism can work in a simple manner: (i) the relevant name is used self-referentially; and (ii) the non-negated predicate in the VP transfers its meaning into the property of NAMING SOMETHING. That is how an utterance of 'Santa doesn't even exist' can be interpreted as (6a).

(6a) 'Santa' doesn't even name something.

---

[12]Geurts 1998 describes both mechanisms in relation with the phenomena of irregular or metalinguistic negations (see Horn 1989).

Presupposition accommodation

Now, on this view, negative existentials involving anaphoric reference would benefit from a different mechanism. The idea is that, just like there is anaphoric reference, there is anaphoric presupposition accommodation by means of which a given assertion may accommodate (or reject) a given presupposition of a previous assertion. In particular, denials may be used to target and reject such presuppositions. Geurts 1998 defends a "binding theory" of presupposition accommodation to explain how presupposition denial works. What I am trying to illustrate is how such mechanism may help hearers interpret a negative existential involving anaphoric reference in a metalinguistic manner. I am not claiming that this mechanism is exactly the one described by Geurts 1998, but rather that there is some mechanism or other of presupposition denial (see Horn 1989). Whatever that mechanism may be should be enough for our purposes.

The "binding theory" of presupposition projection claims that anaphoric reference is just a case of presupposition binding. Just like definite NPs can refer back to objects previously referred to in a given conversation, presupposition carrying expressions refer back to presuppositions of expressions previously used in the discourse. The difference between anaphoric pronouns and presupposition carrying expressions is that the latter may have a richer semantic content. It is this richer content that allows them to be either bound *or* accommodated. On this view, when interpreting the presuppositions of a given use of an expression, speakers are guided by the following constraints in a step-wise manner:

A: Bind the presupposition to a suitable antecedent;

B: If not possible, then accommodate; and

C: If accommodation is necessary, do so globally if not locally.[13]

To illustrate how this mechanism works consider (6b)

(6b) L: Nessie is carnivorous.
     E: No, she isn't. She doesn't even exist.[14]

In (6b) L uses 'Nessie' referentially. So L's use carries a metalinguistic presupposition: *that 'Nessie' has a referent*. This presupposition is inconsistent with E's beliefs. How should

---

[13]To see how this mechanism works in detail and particularly with respect to the principles of Discourse Representation Theory see Geurts 1998, p. 299-304.

[14]Compare with the following case of presupposition denial:

- L: Obama just quit smoking.
  E: No, he didn't. He never did smoke.

she accommodate such presupposition? The metalinguistic view I want to defend takes this presupposition to be accommodated locally within the scope of the negation in E's assertion 'She doesn't even exist'. Here is an example of the kind of reasoning suggested by the metalinguistic view I defend.

E's use of 'she' is anaphoric on L's use of 'Nessie'. It carries the reference and familiarity presuppositions carried by L's use of 'Nessie'. How should one go on interpreting E? Well, if we assume that L's use of 'Nessie' does have a referent we can have a suitable antecedent for 'she'. But doing so would be very costly, for it would make E's assertion inconsistent. We would be taking E to assume there is some such referent of her use of 'she' and that *that* thing does not exist. If we do not assume that L's use of 'Nessie' has a referent then we do not have a suitable antecedent that may bind the reference presupposition of E's use of 'she'. So we are forced to accommodate this presupposition.

Now, we still have two options. We can try to accommodate the presupposition globally, in this case, by whatever appears outside the scope of the negation. If we do so, we would take E to presuppose that her use of 'she' in fact has a referent. But then we are again facing the same problem: E's assertion is inconsistent with this presupposition. Thus, we are forced to accommodate locally, by taking the metalinguistic presupposition to fall under the scope of the negation. If so we can *reject* the presupposition that E's use of 'she' has a referent by taking it to be negated. This delivers a consistent interpretation of E's assertion: it is a denial of L's metalinguistic presupposition *that 'Nessie' has a referent*. Thus, we get once more the desired metalinguistic interpretation that can help the speaker achieve her goal: close off all conversations about Nessie. The resulting interpretation of (6b) is something like (6c).

(6c) L: Nessie is carnivorous.
      E: No, she isn't. 'Nessie' doesn't even have a referent.

5.5.4   Problem solving

Let me now conclude these remarks by showing how well the metalinguistic interpretation can solve the problems posed by negative existentials. These include the use of negative existentials in neutral contexts, affirmative existentials, and existential suppositions.

Neutral contexts

There are two generally different contexts in which a negative existential may be used: charged and neutral. Charged contexts may be said to include or exclude the proposition *that the relevant NP has a referent* among the set of presuppositions. Neutral contexts are

those that may be said to neither include or exclude such a presupposition. I said negative existentials were informative because they convey the information that a given denoting phrase doesn't have a referent. This suggests that they are informative only in charged contexts. However, the account works for both charged and neutral contexts.

Consider a neutral context. We have some evidence of the existence of a big mammal in loch Ness, but the evidence is not good enough to believe there's some such monster. So, we don't know whether or not it exists. In such a context we could utter either one of the sentences in (8).

(8) The Loch Ness monster doesn't exist!
The Loch Ness monster exists!

Asserting either one will be felicitous precisely because the context doesn't make any presuppositions about the reference of the denoting phrase. Asserting either one will also be informative. The negative existential will *convey* that 'the Loch ...' doesn't refer, while the affirmative existential will *assert* that it does. So negative existentials may be felicitously uttered in neutral contexts *and* in contexts that are committed to the existence of a referent of the NP used.

Affirmative existentials

Now, an adequate theory of negative existentials must also explain how affirmative existentials work. For it seems clear that negative existentials are closely related to them. As a matter of fact, a proper account must show how it is that affirmative and negative existentials constitute affirmation / denial pairs. It is a further virtue of the proposal I am defending that it helps clarify this relation. *Affirmative* existentials may be felicitously uttered in both neutral contexts*and* contexts that are committed to there being no referent of the relevant NP. The latter are contexts in which the proposition *that the relevant NP doesn't refer* is being presupposed. The metalinguistic proposal has a very natural way of accounting for affirmative existentials. Consider the conversation in (9)

(9) L: The Loch Ness monster prefers to sleep at the north end of the lake.
E: No, it doesn't. In fact, the Loch Ness monster doesn't exist.
L: You're wrong. The Loch Ness monster exists. I've seen her myself.

Suppose that right after E's assertion the context is negatively charged. So the context is inconsistent with L's beliefs. L can try to to fix the context directly by asserting the affirmative existential. Such an assertion would involve a positive use of the relevant definite NP, so it would carry reference presuppositions. These presuppositions are not yet satisfied by the context, but that doesn't preclude the assertion from being felicitous because, if it

were accepted, the context would have been fixed. The affirmative existential asserts what it presupposes, namely, that a given denoting phrase has a referent.[15]

This is a metalinguistic account of affirmative existentials that explains why affirmative and negative existentials constitute affirmation / denial pairs. Affirmative existentials convey the information that the relevant NP has a referent; negative existentials deny this.

Existential suppositions

I have not said, nor do I want to say, that all negative existential constructions work in exactly the same way. Stalnaker 1978 provides some rather distinct uses of such constructions that will help clarify my position. It seems as if negative existentials act differently in assertions than they do in suppositions. Consider the following counterfactual owed to Kripke:

(10) If Aristotle hadn't existed, the history of philosophy would have been very different from the way it was.

Stalnaker 1978 notes that:

> Clearly the proposition expressed in the antecedent of this conditional is not the proposition that our use of the name *Aristotle* is not appropriately connected with any individual. THAT proposition is compatible with Aristotle's existence. Furthermore, if Aristotle hadn't existed, then our uses of his name probably would not have existed either. The proper name seems to function in the antecedent of the counterfactual more like the way it functions in ordinary predicative statements. The proposition is determined as a function of the PERSON Aristotle; it is true in possible worlds where HE does not exist, and false in possible worlds where HE does exist. [1978, p. 93]

So it seems that negative existential *suppositions* are better understood as involving what I dubbed "positive uses" of the existential construction (see section 5.5.1). In using them, speakers seem to be conveying (or intending to convey) information about the denotation of the relevant NP. If such suppositional uses involve an empty name, the cognitive theory has a ready made explanation of their interpretation. All this is consistent with the metalinguistic account of negative uses that I am defending.

Further issues

It seems to me that there are several other problematic features associated with negative uses of existential constructions. An interesting issue has to do with the fact that it is not correct

---

[15]It is perhaps surprising to say that a given utterance asserts what it presupposes. But there is nothing wrong with this. There are many other expressions the successful use of which takes the speaker to be asserting what she presupposes. Felicitous utterances of 'I am here' and 'I am speaking now' are among them.

to use 'X doesn't exist' when 'X' is the name of a deceased individual. The interesting thing here is that the assertion is incorrect even though it would be literally true and informative. It is not correct, for example, to say things like "Churchill doesn't exist" or "Ted Kennedy doesn't exist". This suggests that the literal interpretation of negative existentials—that which takes assertions of 'X doesn't exist' to convey the literal negation of the claim *that X is an existent thing*—is mistaken.[16]

This feature can be readily explained by the metalinguistic account: if assertions of 'X doesn't exist' convey the information *that the 'X' does not name something*, then an assertion of 'Churchill doesn't exist' would convey the false information that 'Churchill' does not name something. And the same would happen if we were to use any name that is used to refer to no-longer-living individuals within the construction 'X doesn't exist'.

I mentioned above that negative existentials accept NPIs but not PPIs. They have other peculiarities as well. Negative uses, for example, cannot be accommodated prefixally. 'Santa doesn't exist' is not equivalent to 'Santa is unexistent' or 'Santa is inexistent'. Positive uses, however, allow such accommodation. Here are some examples (found online):

(11a) The field of industrial design is inexistent in Lebanon.
(11b) I live in a family where trust is inexistent and suspicion prevails.

Negative uses do not seem to follow the rule of obversion either. 'Santa doesn't exist' is not equivalent to 'Santa is nonexistent'. Positive uses do seem to follow such rule, as in the following examples (found online):

(11c) Service is nonexistent here.
(11d) John's problems are nonexistent.

Negative uses are also not fully compositional. 'Santa doesn't exist' is not equivalent to 'Santa fails to exist'. Positive uses do allow the substitution of 'doesn't exist' for 'fails to exist'. Examples:

(11e) A stable structure may fail to exist even if all the individuals have the same initial endowments in private good [Weber, 1985, p. 178].
(11f) Expected utility fails to exist with constant relative risk aversion [Yoon, 2004, p. 219].

If this evidence is trustworthy, it offers support for the metalinguistic hypothesis, since all these problematic features are, according to the experts (see Horn 1989 and Davis 2008), characteristic of metalinguistic denials. It may also be that these features are not trustworthy symptoms of their being metalinguistic denials. If so, they at least pose a challenge to the opposing view that takes negative existentials to be ordinary negations with a literal

---

[16]The incorrectness of these uses of negative existentials is confirmed cross-linguistically, as it also appears in Spanish and German. This suggests that it is not a peculiarity of some uses of negative existential assertions.

interpretation.

It is unclear how such literal accounts (such as Sainsbury's 2005 and 2010) can explain why negative existentials exhibit all these features: i.e., accepting only NPIs, rejecting prefixal accommodation, violating obversion, and not being fully compositional. Notice as well that on this opposing view the incorrectness in using negative existentials with names of deceased individuals turns out to be a mystery.[17]

5.6   Closing remarks

I have offered a new account according to which competence for empty names demands the use of non-linguistic, general cognitive resources: i.e., something like Leslie's 1987 decoupling mechanism (see section 5.2). On this view, the relevant content associated with ordinary uses of empty names is what I call an EDU (see sections 5.2.3 and 5.2.4). I have shown how appealing to EDUs and their correlate EDU attitudes helps us solve all the puzzles traditionally associated with empty names and a few more having to do with psychological adequacy and the nature of fiction (see section 5.3), as well as some longstanding philosophical puzzles (see section 5.4). I have also shown that this account explains how empty names are interpreted in positive uses of existential constructions, both denials and affirmations (see section 5.5.1). Furthermore, this cognitive theory is compatible with a plausible metalinguistic account of negative uses of existential constructions (see sections 5.5.3 and 5.5.4), which prove to be problematic for any account of empty names. I believe all these are good reasons to accept the cognitive theory of empty names.

---

[17]For arguments against a third alternative account of negative existentials, according to which they are interpreted as quantificational claims, see Walton 2003.

References

Adams, F., and Stecker, R. 1994: "Vacuous singular terms". *Mind and Language*, 9, 4, 387-401.

Baron-Cohen, S., Leslie, A.M., and Frith, U., 1985: "Does the autistic child have a "theory of mind"?". *Cognition*, 21, 37-46.

Bosco, F.M., Friedman, O., and Leslie, A.M. 2006: "Recognition of pretend and real actions in play by 1- and 2-year-olds: Early success and why they fail". *Cognitive Development*, 21, 3-10.

Braun, D. 1993: "Empty names". *Nous*, 27, 449-69

Braun, D. 2005: "Empty names, fictional names, mythical names". *Nous*, 39:4, 596-631

Brentano, F., 1874: *Psychology from an Empirical Standpoint*. London: Routledge and Kegan Paul.

Burge, T. 1974: "Truth and singular terms". *Nous*, 8, 309-25

Byrne, A., 2009: "Experience and content". *The Philosophical Quarterly* 59, 236, 429-51.

Clapp, L., 2008: "The problem of negative existentials doesn't exist". *Journal of Pragmatics* 41,1422-1434.

Davis, W., 2008: "Irregular negations: Implicature and idiom theories" (ms) Georgetown University.

Evans, G., 1981: "Understanding demonstratives". In H. Parret and J. Bouveresse (eds.), 1981: *Meaning and Understanding*. New York: Walter de Gruyter.

Evans, G., 1982: *The Varieties of Reference*. Oxford: Oxford University Press.

Everett, A. 2000: "Referentialism and empty names". In A. Everett and T. Hofweber (eds.) 2000: *Empty Names, Fiction, and the Puzzles of Non-existence*. CSLI: University of Chicago Press.

Everett, A. and Hofweber, T. (eds) 2000: *Empty Names, Fiction, and the Puzzles of Non-Existence*. CSLI : University of Chicago Press.

Fleming, N., and Wolterstorff, N, 1960: "On 'there is' ". *Philosophical Studies* 11, 3, 41-48.

Friedman, O., and Leslie, A.M., 2007: "The conceptual underpinnings of pretense: Pretending is not 'behaving as if' ". *Cognition* 105, 103-24.

García-Ramírez, E., forthcoming a: "An integrated account of substitution failure" (typescript) Philosophy Department, University of Michigan, Ann Arbor.

García-Ramírez, E., and Shatz, M., forthcoming: "On problems with descriptivism: Psychological assumptions and empirical evidence". *Mind & Language.*

Geurts, B. 1998: "The mechanisms of denial". *Language*, 74, 2, 274-307.

Gómez, J.C., 2008: "The evolution of pretence: From intentional availability to intentional non-existence" in *Mind & Language* 23, 5, 586-606.

Heim, I., 1982: "The semantics of definite and indefinite noun phrases" Ph.D. Thesis. Amherst: University of Massachussetts.

Heim, I., 1983b: "File change semantics and the familiarity theory of definiteness". In R. Baurle, C. Schwarze, A. von Stechow (eds.), 1983: *Meaning Use and Interpretation of Language*. Berlin: De Gruyter.

Horn, L., 1985: "Metalinguistic negation and pragmatic ambiguity". *Language* 61, 121-74

Horn, L., 1989: *A Natural History of Negation*. Chicago: University of Chicago Press.

Jacob, P., 2003: "Intentionality". In E. Zalta (ed.) *Stanford Encyclopedia of Philosophy* available online: http://plato. stanford.edu/entries/ intentionality/

Kriegel, U., 2007: "Intentional inexistence and phenomenal intentionality". *Philosophical Perspectives, 21, Philosophy of Mind* 307-40.

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

Leslie, A.M., 1987: "Pretense and representation: the origins of "theory of mind" ". *Psychological Review* 94, 412-26.

Leslie, A.M., 1994: "Pretending and believing: issues in the theory of ToMM". *Cognition* 50, 211-38.

Leslie, A.M., 2002: "Pretense and representation revisited" in N.L. Stein, P.J., Bauer, M., Rabinowitz, (eds) 2002: *Representation, Memory and Development: Essays in Honor of Jean Mandler*. Mahwah, NJ: Lawrence Erlbaum.

Leslie, S.J., 2008: "Generics: Cognition and acquisition". *Philosophical Review* 117, 1, 1-47.

Lillard, A.S., 1993a: "Pretend play skills and the child's theory of mind" in *Child Development* 64, 348-71

Lillard, A.S., 1993b: "Young children's conceptualization of pretense: Action or mental representational state?". *Child Development* 64, 372-86.

Lillard, A.S., 2001: "Pretend play as twin earth: a social-cognitive analysis". *Developmental Review* 21, 495-531.

Lillard, A.S., 2002: "Just through the looking glass: children's understanding of pretense". In R.W. Mitchell, (ed) 2002: *Pretending and imagination in animals and children*. Cambridge: Cambridge University Press

McDowell, J., 1977: "On the sense and reference of a proper name". *Mind* 86, 159-85.

McDowell, J., 1984: "*De re* senses". In C. Wright (ed.) 1984: *Frege: Tradition and Influence*. Oxford: Blackwell.

McDowell, J., 1986: "Singular thought and the extent of inner space". In P. Pettit and J. McDowell (eds.), 1986: *Subject, Thought and Context* Oxford: Oxford University Press.

Nichols, S., and Stich, S. 2000: "A cognitive theory of pretense". *Cognition* 74, 115-47.

Nichols, S., and Stich, S. 2003: *Mindreading: An integrated account of pretense, self-awareness and understanding other minds.* Oxford: Oxford University Press.

Nunberg, G. 1979: "The non-uniqueness of semantic solutions: Polysemy". *Linguistics and Philosophy* 3, 143-84.

Nunberg, G., 1995: "Transfers of meaning". *Journal of Semantics* 12, 109-32.

Nunberg, G., 2004: "The pragmatics of deferred interpretation". In L.R. Horn and G. Ward, 2004: *The Handbook of Pragmatics* Oxford: Blackwell Publishing.

Perner, J., Rendl, B., and Garnham, A., 2007: "Objects of desire, thought, and reality: Problems of anchoring discourse referents in development". *Mind & Language* 22, 5, 475-513.

Perry, J., 1979: "The problem of the essential indexical". *Nous* 13(1), 3-21

Rayo, A., forthcoming a: "An account of possibility" available from http:// web.mit.edu /arayo/www/posex.pdf

Rayo, A., forthcoming b: "An actualist guide to quantifying in" available from http:// web.mit.edu /arayo/www/modality.pdf

Reimer, M. 2001a: "The problem of empty names". *Australasian Journal of Philosophy* 79, 4, 491-506.

Reimer, M. 2001b: "A "Meinongian" solution to a Millian Problem". *American Philosophical Quarterly* 38, 3, 233-48.

Richard, M. 1983: "Direct reference and ascriptions of belief". *Journal of Philosophical Logic* 12, 425-52.

Roberts, C., 1998: "Information structure in discourse: Towards an integrated formal theory of pragmatics". Ohio State University: typescript.

Roberts, C., 2004: "Context in dynamic interpretation". In L.R. Horn and G. Ward, 2004: *The Handbook of Pragmatics* Oxford: Blackwell Publishing.

Russell, B., 1905: "On denoting". *Mind*, 14, 56, 479-93.

Sainsbury, M., 2005: *Reference Without Referents*. Oxford: Oxford University Press.

Sainsbury, M., 2010: *Fiction and Fictionalism*. London: Routledge.

Salmon, N. 1998: "Nonexistence". *Nous*, 32:3, 277-319.

Schiffer, S., 1996: "Language-created language-independent entities". *Philosophical Topics* 24, 1, 149-67.

Searle, J.R., 1975: "The logical status of fictional discourse". *New Literary History* 6, 2, 319-32.

Shatz, M., 1994: *A Toddler's Life*. New York: Oxford University Press.

Soames, S., 2002: *Beyond Rigidity*. Oxford: Oxford University Press.

Stalnaker, R., 1978: "Assertion". In R. Stalnaker, 1999: *Context and Content*. Oxford: Oxford University Press.

Stich, S., and Warfield, T., (eds) 1994: *Mental Representation: A Reader*. Oxford: Blackwell.

Swanson, E., 2006: "Frege's puzzle and pragmatic presupposition". In E. Swanson, 2006: *Interactions with Context.* PhD dissertation. MIT.

Taylor, K. 2000: "Empty names without compromise". In A. Everett and T. Hofweber (eds) 2000: *Empty Names, Fiction, and the Puzzles of Non-Existence*. CSLI Publications: University of Chicago Press.

Thomasson, A.L., 2003: "Speaking of fictional characters". *Dialectica* 57, 2, 207-26

Van der Sandt, R., 1991: "Denial" in *Chicago Linguistic Society*, 27, 2, 331-44.

Walton, K., 1990: *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.

Walton, K., 2000: "Existence as metaphor?". A. Everett and T. Hofweber, (eds) 2000: *Empty Names, Fiction, and the Puzzles of Non-Existence* CSLI Publications: University of Chicago Press.

Walton, K., 2003: "Restricted quantification, negative existentials, and fiction". *Dialectica* 57, 2, 239-42.

Wooley, J., 2002: "Young children's understanding of pretense and other fictional mental states". In R.W. Mitchell (ed.) 2002: *Pretending and Imagination in Animals and Children*. Cambridge, MA: Cambridge University Press.

Weber, S., 1985: "Proportional taxation: Nonexistence of stable structures in an economy with a public good". *Journal of Economic Theory* 35, 178-85.

Yoon, G., 2004: "On the existence of expected utility with CRRA under STUR". *Economics Letters* 83, 219-224.

Chapter 6

Conclusions: Lessons from proper names

What does a philosopher learn by looking at proper names from a cognitively oriented standpoint?

Proper names are cognitively special. They constitute the very first words of an infant's lexicon, their comprehension demands very minimal cognitive resources, yet they are physiologically and memory-wise expensive (see García-Ramírez and Shatz, forthcoming). As we argue in chapter 2, both features appear to have a common source: a lack of semantic associations for proper names. Compared to other items in the lexicon, proper names have very minimal, if any, content. All they seem to demand is an understanding of the reference relation—i.e., that which holds between two objects one of which is taken to be a sign of the other. This understanding appears to be the result of prelinguistic preparedness on the side of the speaker. And it is plausible to think that it is part of a uniquely-human cognitive endowment (see Penn, Holyoak and Povinelli 2008).

Given the plausible assumption that proper names are referential expressions, reference itself turns out to be an abstract yet simple notion. It demands a comprehension of higher order abstract relations. In particular, it requires, first, that the subject understands what it is for an object to be a sign; and, second, that this object can be directly related to a second object of which it is the sign. An understanding of abstract relations such as these appears to be part of a uniquely human cognitive endowment that includes many other abstract notions: e.g., from early on infants understand what it is for an object to be continuous in space and time, as well as what it is for two or three objects to be numerically distinct. Yet, as abstract as it may be, reference is simple: there are no mediating concepts or intensions between the term (or its mental counterpart) and the referent. This yields an important result: human beings do not appear to need mediating intensions to be able to think about objects.

Proper names are also special thought-wise. We normally distinguish between physical or material objects and representations. Cognitive psychologists urge us to distinguish between representations and their interpretations, between, say, a proper name and its assigned referent. We need to make such distinction to understand how it is that humans, as a matter of fact, engage in acts of fiction making (see Leslie 1987). At some point between the second and third year of age human infants are able to follow and *develop* complex games of make-believe that require an understanding of even more abstract notions such as those of "prop" and "representation" (see Nichols and Stich 2000). Thus, signs and, with them, names, are not only good to deliver objects of thought but may themselves constitute relevant guides for thinking. Several different tasks are achieved thanks to this mechanism, including an understanding of pretense, false belief, and plausibly also language processing (see Shatz 2008).

Given a cognitive apparatus that is able to distinguish between a representation and its semantic interpretation, it is not a surprise if that mechanism still distinguishes between representations even when they have exactly the same semantic interpretation. Thus, it is not a surprise that humans may distinguish between 'Without Averroes there would be no western culture' and 'Without Ibn Rushd there would be no western culture'. Furthermore, it is not a surprise that competent speakers may form their opinions based on this distinction. In chapter 3, I argue that making such distinctions and forming opinions on their basis are intimately related phenomena. Without an understanding of the underlying cognitive mechanisms, such mental phenomena become puzzles: "informativeness" and "substitution failure".

If one is merely concerned with assigning a semantic interpretation (e.g., reference, content, truth-conditions) it will of course seem puzzling that a substitution of semantically equivalent representations should make any difference. And it may seem natural to think that one needs to add something else to the semantics. The picture of the human cognitive apparatus offered in chapters 2 and 3 suggests that human cognition is far from being only interested in the semantically assigned interpretation of the representations toward which subjects may direct their thoughts.

In chapter 4, I argue that the key to solve these puzzles is to realize that the relevant thoughts are concerned with the representations and not so much with their semantics. Subjects may take different coreferential names to refer to different objects and, hence, understand the predicates associated with each *name* as corresponding to different objects. Speakers may realize this and exploit it in order to adequately describe the mental state of such subjects. The clue, then, is not to make the semantics more complex but to find out what is guiding the relevant thought, that is, to realize what the relevant thought is *really*

about. I take this to be one of the advantages of the cognitively oriented methodology I have followed throughout (see Leslie 2008 for a similar approach).

Finally, proper names are special as they are unique to our human way of living. Understanding names illuminates how we think about how our world is, but also how we imagine how it could be and should be: i.e., how we create fictional stories as well as scientific hypothesis. It seems natural to think that the cognitive mechanisms involved in human pretense and fiction making would play a significant role in the way in which humans use fictional names (see Shatz 2008). Cognitive psychologists have done a lot to improve our understanding of such mechanisms.

There is evidence suggesting that humans make use of something like a decoupling mechanism by means of which they copy, manipulate, and interpret mental representations according to the demands of the situation. Some such mechanism has been shown to play a central role in the comprehension and ascription of mental states, as well as in the comprehension and engagement with pretense and make-believe (see Leslie 1987, and Friedman and Leslie 2007).

In chapter 5, I take seriously the idea that there is a common mechanism underlying acts of pretense and the understanding of empty names. Thus, I offer a cognitive theory of the ordinary use of empty names by appealing to the decoupling mechanism that psychologists associate with pretense. The account is offered not in terms of the semantics of empty names, which I take to be the same as that of non-empty names, but in terms of cognitive-general, non-linguistic, abilities of interpretation and manipulation of representations. I develop the account further by including a somewhat detailed description of the type of representation that is produced by such psychological processes. Once we have such a "cognitive" content in hand we can explain how humans manage to convey information, form opinions, and construct complete and meaningful stories, as well as theories, by using empty names.

There appears to be more than just fiction and science to human existence. We also live in a world replete with institutions and norms. The cognitive theory, I believe, can be extended to contribute, at least partly, to a better understanding of how we have thoughts and convey information about such things. As things stand now, there are good reasons to think that moral cognitive development is essentially tied to the development of a theory of mind, both of which seem to make use of the same decoupling mechanism deployed in acts of fiction making (see Shatz 1994, Wellman 1990, Haidt 2008, and Wellman and Miller 2008). It is not a surprise, on this view, that some autistic children with poor moral cognition tend to do bad in false-belief tasks and fail to engage in acts of make-believe.

We can learn a lot by looking at proper names from a cognitively oriented standpoint.

≪ – Et ces't tout?

– Ça suffit!≫

— Antoine de Saint-Exupéry

Le Petite Prince

References

Friedman, O., and Leslie, A.M., 2007: "The conceptual underpinnings of pretense: Pretending is not 'behaving as if' ". *Cognition* 105, 103-24.

García-Ramírez, E., and Shatz, M., forthcoming: "On problems with descriptivism: Psychological assumptions and empirical evidence". *Mind & Language.*

Haidt, J., 2008: "Morality". *Perspectives on Psychological Science* 3, 1, 65-72.

Penn, D.C., Holyoak, K.J., and Povinelli, D.J. 2008: "Darwin's mistake: explaining the discontinuity between human and nonhuman minds". *Behavioral and Brain Sciences.* Cambridge: Cambridge University Press.

Leslie, A.M., 1987: "Pretense and representation: the origins of "theory of mind" ". *Psychological Review* 94, 412-26.

Leslie, S.J., 2008: "Generics: Cognition and acquisition". *Philosophical Review* 117, 1, 1-47.

Nichols, S., and Stich, S. 2000: "A cognitive theory of pretense". *Cognition* 74, 115-47.

Shatz, M. 1994: *A Toddler's Life*. New York: Oxford University Press.

Shatz, M. 2008: "Language as a consequence and an enabler of the exercise of higher-order relational capabilities: Evidence from toddlers". *Behavioral and Brain Sciences* 31, 145-46.

Wellman, H.M., 1990: *The Child's Theory of Mind*. Cambridge, MA: Bradford, MIT Press.

Wellman, H.M., and Miller, J. G., 2008: "Including deontic reasoning as fundamental to theory of mind". *Human Development* 51, 105-35.