EDUARDO GARCÍA-RAMÍREZ

OPEN COMPOSITIONALITY

TOWARD A NEW METHODOLOGY OF LANGUAGE

Open Compositionality

Open Compositionality

Toward a New Methodology of Language

Eduardo García-Ramírez

LEXINGTON BOOKS
Lanham • Boulder • New York • London

Published by Lexington Books
An imprint of The Rowman & Littlefield Publishing Group, Inc.
4501 Forbes Boulevard, Suite 200, Lanham, Maryland 20706
www.rowman.com

6 Tinworth Street, London SE11 5AL, United Kingdom

British Library Cataloguing in Publication Information Available

Library of Congress Cataloging-in-Publication Data

Names: Garcia Ramirez, Eduardo, author.
Title: Open compositionality : towards a new methodology of language / Eduardo Garcia-Ramirez.
Description: Lanham : Lexington Books, 2019. | Includes bibliographical references and index.
Identifiers: LCCN 2019018976| ISBN 9781498562720 (cloth : alk. paper) | ISBN 9781498562737 (electronic)
Subjects: LCSH: Compositionality (Linguistics)
Classification: LCC P325.5.C626 G37 2019 | DDC 401/.43—dc23
LC record available at https://lccn.loc.gov/2019018976

™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI/NISO Z39.48-1992.

Printed in the United States of America

To Florencia and Gastón

Contents

Preface

According to Dummett (1973), as theorists interested in natural language,

> 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, 92)

This book is an attempt at following Dummett's imperative unreservedly. Traditionally, philosophers of language have endorsed a strongly compositional account according to which humans understand natural languages in virtue of possessing specialized cognitive machinery—made up of a lexicon, syntax, and semantics—and that natural language works by assigning the meaning of any complex expression by means of an algorithmic compositional function. I am convinced that this view—here dubbed the "closed view"—is mistaken. To better understand how it is that humans actually understand natural languages and, just as importantly, to find out what is it that they understand, I look into empirical studies on language acquisition and development, as well as studies on language processing.

The result is not a welcoming one for the closed view. There is no specialized cognitive machinery by means of which humans acquire and understand natural languages and, unsurprisingly, natural languages turn out *not* to be strongly compositional systems of representation, the job of which is to algorithmically assign meanings to complex expressions. Instead, natural languages appear to be the result of a higher order, supermodular cognitive architecture constantly interacting with multiple domains and cognitive

mechanisms, far more than those traditionally considered linguistic (i.e., lexicon, syntax, and semantics). So conceived, natural languages are not just in the business of assigning meanings to complex expressions—something they do by multiple different means, in many cases without recurring to any sort of compositional or syntactically driven process—they are also in the business of shaping human cognition by offering a highly interactive thinking platform. This is the view of natural language that I defend in this book. I call it "open compositionality."

Chapter 1 sets the stage by presenting two different views of language. The traditional view in the analytic tradition, championed by Montague in the 70s and followed by so many philosophers of language and linguists nowadays, is presented as the closed view. According to the closed view natural languages are systems of representation with the goal of communication, linguistic practice is in the business of communicating thoughts among participants, and all this is possible thanks to the compositional nature of language, which determines meanings by means of syntactically driven algorithms. The other view, better known within cognitive psychology and psycholinguistics, is presented as the open view. According to the open view natural languages are complex, higher order cognitive capacities playing a central role in every individual's cognitive development. Language is, on this view, a powerful platform for developing human unique cognition. I argue that both views are complementary, showing how we can only benefit from a successful synthesis of both. To do so I propose to look into the closed view and its theoretical commitments to determine which elements of the theory may be kept, in virtue of their consistency with the empirical evidence, and which must be abandoned.

Chapter 2 offers an account of the closed view. I offer a brief historical account of the view in order to highlight its motivation and closeness to the development of mathematical logic. The central assumption of the closed view is, of course, the principle of compositionality, which I discuss at length in this chapter. Following Szabó (2012), I present multiple distinct versions of the principle, arguing that only a *strong*, logically closed version is to be seriously considered. I describe the methodology associated to the principle of *strong compositionality* and show how simple, clear, and easily generalizable it is, making it an extraordinary methodology of language. I present three well-known arguments intended to offer independent empirical support for the closed view—*productivity, systematicity,* and *computability*—and conclude by arguing that, explanatory power aside, the compositional methodology runs the risk of being trivial unless it works under substantial empirical assumptions.

Chapter 3 considers the empirical commitments of *strong compositionality* in order to test them against the evidence. This chapter plays a central role

in the general argument as it presents a complex and detailed empirical argument against the idea that natural languages observe *strong compositionality* and, thus, against the closed view. I show that *strong compositionality* carries two substantial commitments concerning the specificity and cognitive architecture of linguistic cognition, and a third one concerning the algorithmic nature of language processing. I consider empirical evidence on language acquisition and development, the acquisition of words and syntactic knowledge, and the very nature of language processing among adult speakers. The evidence shows that all three commitments of *strong compositionality* are empirically false. Natural languages are not *strongly compositional* systems of representation.

Chapter 4 presents an alternative, novel view of language and linguistic practice based on the evidence considered in chapter 3. With respect to knowledge of language the proposed view, which I call "open compositionality," considers language to be a supermodular cognitive ability, capable of benefiting from the interaction with multiple domains and cognitive mechanisms. With respect to linguistic practice, the view offers an account of semantics as a decision-making process. As a result, the meaning of any complex expression turns out to be the most *appropriate*—given practical contextual limitations—and *economic*—given time and mental resources available—*interpretation*—given limited information available. To find the most appropriate and economic interpretation speakers need not always follow syntactically driven interpretations, given the interactive nature of language, they may also benefit from heuristic strategies as an aid for achieving fast and frugal decision-making. Aside from this unorthodox semantics, open compositionality also comes with a novel "cognition first" methodology, according to which in order to account for any problematic linguistic phenomenon we must first earn a clear understanding of the cognitive processes underlying it. Chapter 4 concludes by showing how open compositionality explains the *productivity, systematicity*, and *computability* of natural languages just as well as, if not better than, *strong compositionality*.

Chapters 5, 6, and 7 are meant to exhibit the explanatory power of open compositionality and the cognition first methodology by showing how the proposed account can be applied to obtain successful accounts of three traditionally problematic, recalcitrant phenomena. Chapter 5 looks into the phenomena of substitution failure and offers an account in terms of lexical processing architecture. Chapter 6, previously published as a separate article, is concerned with empty names and their meaningful uses. The account offered claims that speakers make use of a decoupling mechanism allowing them to divorce representations from their content and arbitrarily assign any alternative content as surrogate. Chapter 7 considers the case of moral discourse and the meaning of moral terms from the point of view of a naturalist

moral realist. It is argued that there is a substantial analogy between language and morality, between knowledge of language and moral knowledge. Following cognitive psychologists and primatologists (see Hamlin 2013; and de Waal 1991) an outstanding evolutionarily endowed cognitive capacity for moral thinking, our "moral competence," is postulated. It is shown how this competence can account for human moral knowledge, moral motivation, moral objectivity, and, of course, for how moral discourse and moral terms get to be interpreted, as language processing can benefit from moral competence to process moral representations. This competence-based account is, of course, little more than a sketch, but it is a promising, novel alternative for those interested in moral knowledge from a naturalist perspective.

Having said what this book is about it will be useful to clarify what this book *is not* about. First and foremost, this book is not intended to defend a holistic account of language, or anything closed to that. Open compositionality is not a form of semantic holism or conceptual-role semantics. Open compositionality is, after all, a compositional view of natural language and, as such, it presupposes the possibility of having an individual assignment of word meanings. Briefly put, open compositionality departs from *strong compositionality*, but it is not tantamount to *non-compositionality*. Its departure from the traditional endorsement of *strong compositionality* merely consists in the rejection of its logical closure. Open compositionality accepts that the meaning of complex expressions may be fully determined by syntactically driven algorithms, it simply rejects that this is how the meaning of all possible complex expressions, and even all uses of one and the same complex expression, must be determined. In so doing open compositionality allows for alternative meaning determination procedures, like the use of fast and frugal heuristic strategies. The difference is minimal yet enormously consequential.

Second, this book does *not* pretend to offer a theory of child language, as if open compositionality were compatible with a *strongly compositional* view of *adult* language. Open compositionality is based on studies on language acquisition and development involving young infants. But that is only part of the evidence considered, and it is so used only as it is generally considered to be evidence of the very nature of human languages. The proposal is also based on studies on adult knowledge of language and how this knowledge is used for language processing. This book is *not* about knowledge of language as an end in and of itself, but about human languages themselves. It is assumed that the very nature of the latter is closely tied to human knowledge of language. If the reader is irremediably convinced that a philosophical account of language has absolutely nothing to do with human knowledge of language, this might be the farthest she goes on reading this book. Instead of continuing, the reader might want to ask herself what kind of thing is a

natural language if it is not, for the most part, something in the head of, and the interactions in between, those speaking it.

The title of this book, and of the theory proposed, is both a description and an imperative. It describes a view of natural languages according to which they are both open *and* compositional. They are *open* because they allow for different interpretation procedures, including various heuristic strategies, of meaning determination. As such, the set of possible meanings for any given complex expression is completely open. It allows for ranks, ties, and conflicts among alternative interpretations; and it may also be that one and the same expression may have different meanings under different practical, economical, and informational conditions. Yet, natural languages are still compositional, as they allow for syntactically driven linguistic processing to be always present although playing different roles depending on the relevant practical, economical, and informational limitations. As I said, the title is also meant as an imperative or, perhaps better, a suggestion to my peers. ¡Open compositionality!

In a sense I have been writing this book for the past nine years. I began working on some of the ideas of the book in the spring of 2010, after defending my PhD dissertation at University of Michigan. Back then I couldn't know that I was moving closer to open compositionality. I now think that even my dissertation is permeated with it. Nine years are quite a bunch. The list of people that I have met and discussed these ideas with is too long to be exhausted (or even remembered). Special thanks are owed to Marilyn Shatz, my mentor, for her patience and guidance, and also for encouraging me to take long enough time to reflect on these ideas and turn them into a book. I hope nine years were enough. I also want to thank Florencia Rimoldi, my life partner, for her unwavering support and understanding, for emboldening me to go on with this difficult project, for helping me find the time to write, and, last but not least, for her careful and thorough reading of the manuscript. I also want to thank my son, Gastón, for letting me work from 4 to 7 everyday, with no exceptions, and then offering me a live seminar on how to be human. Especial thanks are also owed to an anonymous referee for her poignant reading and constructive, reassuring comments.

Through the years the ideas in this book have greatly benefited from discussions with colleagues at home and abroad. I am greatly thankful to Axel Barceló, Maite Ezcurdia, Daniel Drucker, Mario Gómez-Torrente, and Ricardo Mena. Thanks also to the members of the Philosophy of Language Seminar at the Instituto de Investigaciones Filosóficas, UNAM, for taking the time to read previous versions of the manuscript. Different parts of this book were presented at various workshops and conferences. I am thankful to audiences at the SPR-11 Workshop (Donostia, Basque Country); the Current Projects Seminar Series of the Philosophy Department, University of Sydney (Sydney,

Australia); the 1st Workshop on Reference at IIFs-UNAM (México); the Philosophy Department at Karl-Franzens Universität (Graz, Austria); the Philosophy Department at Ruhr Universität Bochum (Bochum, Germany); the 4th Workshop on Philosophy and Cognitive Science—SADAF (Buenos Aires, Argentina); the 3rd International Meeting of the Argentinian Association for Behavioral Sciences-Universidad Nacional de Córdoba (Córdoba, Argentina); the 5th meeting of the Language, Context, and Cognition Workshop-Universidad de los Andes (Boyacá, Colombia); and the 4th meeting of the ICSO-Workshop—SADAF (Buenos Aires, Argentina). Especial thanks are owed to Rob Stainton, Lisa Skydelsky, and Eleonora Orlando. I also want to thank the members of the BA-LingPhil Group (Buenos Aires, Argentina) for letting me be part of such a great research group while I was working on this book. Thanks are also owed to Miguel Gama and Pedro Espinosa, at the Eduardo Garcia-Maynez Research Library, for granting me access to all the research material required for this project, both at home and abroad. Finally, I want to thank my students from various graduate seminars, both in Mexico (UNAM) and Argentina (UBA), for not letting anything pass without rigorous scrutiny. Thanks are owed to Analía Zilber, Aframir Montero, Axel Fernández, Erika Torres, Fernando Carranza, Gisela Martínez, Julieta Straccia, Laura Campos Millán, Micaela Difalcis, Natalia Curtidor, Nicolás Serrano, Ramiro Caso, and Romina Trebisacce

The research for this book were supported by the research project PAPIIT-IN400915. I also benefited from a sabbatical research grant from PASPA-DGAPA; a sabbatical research grant from CONACYT; and a yearlong sabbatical leave from my home institution, IIFs-UNAM. Thanks are owed to the Sociedad Argentina de Análisis Filosófico (SADAF) for hosting me during this period.

Chapter 1

Two Views of Language

What is a language? David Lewis (1975) famously defends a double answer to this question. Lewis' goal is to show that two seemingly incompatible views are in fact consistent with each other and even complementary. On the one hand, there is a formal conception of language inaugurated by Frege (1892) and Russell (1905) and substantially developed by Montague (1974). According to this approach languages are abstract entities, more specifically, they are sets of ordered pairs of sentences and meanings. The goal of a philosophical theory of language, from this perspective, is to account for the function that delivers the correct set of such ordered pairs for a given language. On the other hand, there is a social conception of language, commonly associated to Wittgenstein (1967), Quine (1960), Austin (1962), and Grice (1989). According to this approach a language is a social phenomenon that takes place within the sphere of human actions. The goal of a philosophical theory of language, from this perspective, is to account for the multiple goals, and means to achieve those goals, that speakers may have while using a given language.

Lewis (1975) argues, successfully I think, that the formal and social approaches to language may, and should, be synthesized into a single more complete account. Lewis (1975) offers an account of the social aspect of human languages in terms of his proposed analysis of convention and, hence, takes human languages to be abstract sets of ordered pairs of sentences and meanings that happen to be used by the members of a given population to achieve their goals. The key of the account, claims Lewis, is to understand that such use of an abstract object is governed by social conventions, in particular, by the conventions of truthfulness and trust in the said language.

Yet, independently of the success of the synthesis, both the formal and social approaches to language are part of a single view, perhaps the most

widespread view within the philosophy of language in the analytic tradition. According to this view, whether formal or social, a language is a compositionally closed representational system useful for communication. This view of language is certainly not the only approach in the literature. Since at least Kant (1999), the philosophical tradition has also considered language from a rather different perspective. According to this alternative approach, human language is a cognitive tool useful for knowledge acquisition and cognitive development. Within philosophy, this view of language developed both inside (see Quine 1960; Wittgenstein 1967; Brandom 1994; and Sellars 1997; among others) and outside the analytic tradition (see, for example, Heidegger 1962; Gadamer 1989; and Foucault 1972).

Arguably, the most advanced and better-supported development of the latter approach has flourished beyond philosophy. Within empirical psychology this view of language has thrived, as it accompanies the advancement of cognitive and developmental psychology, as well as psycholinguistics. Decades of empirical studies on language acquisition and development (for a recent overview, see Hoff and Shatz 2007) have substantiated the claim that—just as it happens with other cognitive abilities such as object tracking, mind reading, and mathematical thinking—there are specific conditions for language acquisition, while language development exhibits certain predictable patterns. According to this view, more than just a compositional system of representation, human language is a cognitive ability and a rather special one, as it appears to be open to the interaction with, and modification from, the exercise of higher order capabilities (see Penn, Holyoak, and Povinelli 2008; Shatz 2008).

Both views of language have proven to be a great theoretical success and may, at first glance, seem fully complementary. By endorsing the principle of compositionality, the closed view has extended its understanding of natural language semantics to include theories of proper names, definite descriptions, demonstratives, indexicals, adjectives, predicates, vague expressions, modal expressions, speech acts, implicatures, and presuppositions, among other phenomena. Yet, the open view has substantially improved our understanding of language acquisition and development allowing us to know when, how, and why are certain kinds of expressions acquired; why proper names are among the first words in an infants' lexicon; how syntax, semantics, and pragmatics interact in development; and why property concepts and terms are comparatively difficult to learn, among many other things. However, these views of language have been kept apart more often than not, and more so in philosophy of language than in cognitive psychology and psycholinguistics.

In this book I argue for a distinct view within philosophy of language, one that considers elements of both views conceiving natural language as both

systems of representation as well as powerful higher order cognitive tools.[1] My goal is to describe an alternative philosophical view, according to which natural language is compositional yet openly so. The resulting view, according to which natural language is more than just a representational system and that, as such it is not compositionally closed, offers a more satisfactory account of philosophically interesting phenomena—e.g., substitution failure, empty names, and moral discourse—while at the same time being consistent with (as it benefits from) our best theories of language acquisition and development, cognitive development, pretense understanding, and language processing, among others. I will call this "open compositionality" and will develop it in chapter 4.

1.1 THE CLOSED VIEW: CLOSED COMPOSITIONALITY

What is a language? According to the closed view, human languages are complex communicative tools, the use of which give place to several important questions. How does linguistic practice take place? What is the relation between what we say, what we think, and what we do? How does what we say to others relate to what they reply to us? The analytic tradition in philosophy of language has aimed at answering these questions by following a methodology based on the logical analysis of language, especially the logical analysis of ordinary language use, and with the help of two central hypotheses.

According to the first hypothesis competent speakers use language to *express* their thoughts; ordinary language use somehow reflects our mental states. Similarly, what speakers say in reply to what others have said is taken to reflect what the former think about the issues at hand. This relation between what we say and what we think is also meant to illuminate the relation between what we say, what we think, and what we do. This latter relation comes at different levels of complexity. There is an intuitive level at which the relation is understood in terms of a correspondence between the content of our thoughts and the meaning or content of our words. This correspondence can be understood as a set of semantic relations (at least partly) determining the meaning of what we say. But there is also a less intuitive level at which the relation is the product of multiple different phenomena that account for why a given population uses this or that particular language and not another one—why some populations use English, for example, and not Spanish or German—and why they use it in the way they do, by following these or those rules or maxims and not different ones. These complex phenomena can be understood, following Lewis (1975), as a set of regularities in action and / or belief that give place to the social aspect of language use. Together the semantic relations and the regularities in action and / or belief are meant to

explain how the meaning of the complex expressions we use (e.g., sentences, phrases, etc.) corresponds to the content of our beliefs, desires, fears, etc.

According to the second hypothesis of the closed view, competent use of language does not only follow certain rules of correspondence and social conventions, it also presupposes (if it does not directly exhibit) a specific logical structure that results from observing a particular set of rules of composition. These rules of composition, it is assumed, *fully determine* the meaning of any complex expression of the relevant language. Knowing such rules, it is assumed, is part of what it takes to be a competent speaker. It is by means of them that a speaker may understand and identify the meaning of *any* complex expression in her language, thus allowing her to recognize the sentence of her language that best corresponds with the thought she wants to express, as well as the meaning of whatever expression others may have used to express thoughts of their own.

Among others, these rules include syntactic/semantic rules of combination and, most importantly, they are governed by the principle of compositionality. Since these rules are taken to *fully determine* the meaning of *any* complex expression in a given language, one can say that, on this view, human languages are *closed under compositionality*. Briefly put, the hypothesis claims that all complex meanings—i.e., the meaning of all complex expressions—are compositionally determined (for a more detailed account see Szabó 2013; Werning, Hinzen, and Machery 2012). As we will see, both in this section and with more detail in chapter 2, this hypothesis offers a principle that is *closed* in what turns out to be an important *logical* sense, namely, by having all its quantifiers be universal in scope. It is a hypothesis that is meant to apply to *all* possible complex expressions and is meant to determine *all* possible meanings for such expressions. Such *closure* is intended to preclude the possibility of there being non-compositional processes (whichever they may be) involved in the determination of meaning for complex expressions. In other words, according to the second hypothesis of the closed view, if a given sentence *S* is made up of simple expressions $<e_1, e_2, \ldots e_n>$ combined in way *W*, and the principle of compositionality determines that, when combined in way *W*, $<e_1, e_2, \ldots e_n>$ mean *M*, then *S* means *M* and nothing else. This hypothesis of compositionality, or of compositional closure, plays a central role in the closed view of language in the analytic tradition, in particular it has become a defining feature of what is nowadays the orthodoxy in philosophy of language.

This hypothesis of closure under compositionality is best expressed by the principle of compositionality. There are multiple versions of this principle differing on their target notion—i.e., content, truth-conditions, informativeness, syntactic structure, or even a syntax-to-semantics determination

relation. For now it will be enough to consider a basic version of the principle, such as the following.

Compositionality: The meaning of any complex expression *E*, made up of the simple expressions $<e_1, e_2, \ldots e_n>$, of a language £ is fully determined by the meaning of $<e_1, e_2, \ldots e_n>$ in £, and the way these are combined in *E*.

This principle embodies several of the distinctive features of the closed view of language in contemporary philosophy. First, according to this principle the meaning of a complex expression is an *intrinsic* property of such linguistic expression. In other words, the meaning of a complex expression *E* is determined as a property that *E* itself has in virtue of what it is, and not in virtue of the relations it holds with other expressions, or with the intentions or presuppositions of the speaker/hearer, or of the way the world is independently of *E*. There are, according to the principle, two necessary conditions that, together, are enough to determine the meaning of *E*. These are, first, the meaning of the simple expressions $<e_1, e_2, \ldots e_n>$; and, second, the way they are combined in *E*. Both conditions are part of what makes *E* what it is, independently of any other expression; even more, they are part of what makes *E* the expression it happens to be independently of how *E* is used by any competent speaker.

To illustrate consider the complex expression "The kids are playing in the snow." According to the principle, this expression's meaning is determined, first, by the meaning of "The," "kids," "are," "playing," "in," and "snow" in English. Each one of these expressions lacks (at least at first glance) any internal structure, so they can be considered to be simple expressions of English. As such, each one may be said to have an assigned meaning in the English lexicon, or in the speaker's vocabulary. Thus far there is no need to look for anything beyond the complex expression itself, it has "The," kids," "are," "playing," "in," and "snow" as its simple constitutive expressions, and they in turn have assigned meanings (according to the relevant lexicon). Furthermore, the syntactic relations that hold among these simple expressions determine the way in which they are combined. The latter, according to the principle, is the second and last requirement needed to determine the meaning of the sentence. There is no need to look beyond the complex expression itself to determine either the meaning of the simple expressions or the way in which they are combined. Thus, the meaning of a complex expression is, according to the principle of compositionality, an intrinsic property of it.

A second important feature of the principle is the *determination relation* that is said to hold between the simple expressions, the way of combination, and the meaning of the complex expression. This relation is commonly

understood as a mathematical (recursive) function that, given the values (i.e., meanings) of the simple expressions $<e_1, e_2, \ldots e_n>$ in £ delivers the value of the whole complex expression *E* in £. In doing this, the closed view has gained a lot in terms of explanatory power. If language is in fact compositionally closed and if compositionality can be fully explained in terms of mathematical functions, then language becomes a much simpler, tractable phenomenon. Natural languages, such as English, do not differ substantially from formal languages, such as propositional logic, on this view. To fully explain a given natural language £ all we need to do, as theorists, is to identify the proper mathematical functions. This should be rather straightforward once we are given the lexicon and the syntax of £ though, of course, getting the lexicon and (most importantly) the syntax correctly is not an easy task and may require a lot of work (both empirical and formal).

By giving place to such an understanding of natural language, the closed view can, at least in principle, be easily extended to offer a semantic account of linguistic practice in general. There are only three variables a theorist must identify, the meaning of simple expressions, the way they are combined (which depends on the syntactic structure of the expression), and the syntactic-semantic rules that govern such combinations. Given that this is a *closed* system, with no extra elements playing a role, complete semantic accounts are straightforward. A simple methodology suggests itself. Whenever a theorist has trouble identifying any of the three necessary elements, she can infer it (almost by *subtraction*) by fixating on the remaining two. Competent speakers know, in virtue of their competence, what the meaning is of any complex expression, which are the syntactic-semantic rules of combination, and are usually reliable when it comes to identifying the way in which simple expressions combine to form a complex one.

If, for example, our theorist is trying to offer a semantic account of proper names, all she needs to do is identify as many distinct complex expressions involving proper names as possible, fix on a proper interpretation (i.e., meaning) of such expressions, and identify the syntax of such expressions and the place that proper names occupy in them. The correct semantics for proper names will simply be—on this view, of course—whatever it is that can explain how we get from the identified syntax of the various expressions to their proper interpretation. Sometimes this methodology requires further intermediate steps or some negotiation between a simple straightforward semantics and a more complex syntax (see, for example, Matushansky 2006 and 2008).

Such a transparent and simple methodology has proven to be one of the most important achievements of the closed view. It is as easy to understand as it is to follow, thus becoming a rather efficient and widespread theoretical tool. Furthermore, its acceptance of distinct moving parts—e.g, the meaning

of simple expressions, the syntax of the complex expression, and the way of combination—allows for rich and substantial debates not only about the semantics of the target expressions, but also about the detailed structure of the communicative view itself (see the essays in Szabó 2005; see also Ezcurdia and Stainton 2013).

Finally, aside from having a simple and clear explanation of how complex expressions get their meaning, as well as an efficient methodology with an all-inclusive formal scope, by endorsing the hypothesis of compositionality the closed view also delivers a straightforward story about linguistic *competence* and, thereby, linguistic practice. According to this view to be a competent speaker of £ is to know the relevant parts and structure of the corresponding compositionally closed system; i.e., to know the meaning for simple expressions in £ (i.e., £'s lexicon), the rules of formulation in £ (i.e., £'s syntax), and the rules for functional content (i.e., £'s compositional semantics). Given how complex linguistic expressions may be, and how limitless human languages are in terms of expressive power, this is a surprisingly simple and elegant account. To explain how speakers of £ may understand all kinds of expressions of it, from simple phrases to full texts and an unlimited number of them, all we need is these three ingredients. It may be difficult to offer all the details of the relevant lexicon, syntax or compositional semantics, but it is certainly achievable. Furthermore, it is no mystery why, on this view, human beings are so good at learning human languages. After all, they certainly have enough cognitive capacities to learn a finite lexicon, as well as a finite set of syntactic and semantic rules. Aside from what may be needed to acquire the lexicon, no *extra-linguistic* or *non-linguistic* knowledge is required.

It is not hard to see how linguistic practice takes place from this point of view. In virtue of her knowledge of the lexicon, syntax, and compositional semantics of £ a competent speaker can determine the meaning of all, simple and complex, expressions of £. If so, then such competent speaker will be capable of determining which (simple or complex) expression of £ has a content that corresponds with that of her thoughts. Likewise, she will be able to determine the content of other speakers' thoughts by taking them to correspond with the content of the (simple or complex) expressions they use.

This should be enough for an initial approach to the closed view. In chapter 2, I will offer a more detailed account of it, explaining how it is meant to work and why it is such a successful account. For now it is important to focus on three features of the view. First, it considers natural language to be essentially a tool for communication. Human languages are, on this view, in the business of forming meaningful representations and only that. Second, such communicative tool has a very specific formal nature, i.e., it is compositionally closed. Human languages are made up of a lexicon, a syntax, and a compositional semantics, all of which correspond to a very specific domain as well, as they

are specifically about language processing. Third, these two features suggest an understanding of natural language as an autonomous cognitive system, one that works independently from other human cognitive abilities; all that this communicative system needs to work properly is a set of purely linguistic tools (lexicon, syntax, and compositional semantics).

1.2 THE OPEN VIEW

What is a language? According to a more cognitive approach, natural language is a higher order cognitive ability that all human beings naturally (barring abnormal cases) exercise competently. From this perspective, the focus is put on the cognitive resources needed to achieve such competence and the process by means of which a prelinguistic infant becomes an expert language user as an adult. As such, this approach must face multiple different questions concerning language acquisition and development (see Bloom 1993). Is it acquired or does it just develop? Is the stimulus for acquiring a language really poor? Is language learning an autonomous enterprise or is it a product of the interaction between multiple cognitive abilities? Are there only linguistic constraints or are there multiple kinds of constraints? Is linguistic competence achieved instantaneously, with limitations having to do merely with performance, or is it the result of a stage-like developmental process? For over a hundred years now, these questions have been empirically addressed by means of ever more detailed and careful experimental designs going from diaries tracking specific subjects through their infancy all the way to the use of fMRI's, PET scans, and eye-tracking devices on multiple subjects (for a recent review of the field, see Hoff and Shatz 2007).

This cognitive approach works under the assumption that understanding a subject *S*'s knowledge of a language £ is essential to understanding £ itself. This is so because, arguably, the very nature of the language £ that a speaker *S* possesses is determined by *S*'s knowledge of £ (see Chomsky 1986). The study of language acquisition and development just is the study of what human beings know when they know a language—i.e., phonology, syntax, lexicon, semantics, pragmatics, social knowledge, etc.—how they get to know it—i.e., which learning mechanisms are employed and under which constraints—and how is such knowledge used from infancy into adulthood—i.e., how does knowledge of language and, consequently, how does language itself evolve through the lifespan.

Chomsky (1965) proposes a very influential account of language acquisition according to which it happens in virtue of a universal grammar. A universal grammar is some kind of language acquisition device (mainly for syntax) that carries constraints on what possible languages there could be and

specifications of what they might consist of. These constraints, or principles of universal grammar, include both syntactic and semantic rules (see Larson and Segal 1995). The universal grammar is meant to be domain specific and autonomous. It is domain specific because it only includes linguistic principles, that is, principles having to do either with syntax, semantics, or both. Thus the learning, or otherwise cognitive, mechanisms it demands are expected to be of use only for language learning. It is autonomous because its components (i.e., syntax and semantics) are not sensitive to the influence of other, domain general, principles or learning mechanisms. As such, the universal grammar can output any and only grammatical sentences by itself. On this approach language acquisition and development is little more than the process of maturation of the universal grammar.

Departing from the Chomskyan proposal, the open view of language is based on two central tenets. First, it claims that language is a product of a process of acquisition characterized by the interaction between multiple domain general—i.e., not language-specific—learning mechanisms, such as statistical analysis, distributional information analysis, understanding of mental states, and tracking predictable patterns, among others. Second, the view also claims that language is subject to developmental processes of change, and not just maturation.

Most recent studies in early language acquisition and language development support both tenets of the open view. There is mounting evidence suggesting that language acquisition is not instantaneous but subject to development, that it demands the use of multiple domain general learning mechanisms, and that the components of language (i.e., syntax, semantics, pragmatics, phonetics, etc.) are not autonomous. Word learning, for example, is a very difficult task. The input is never transparent, and hearers must process multiple inputs, consider many alternative possible meanings, run a statistical and probabilistic analysis, and narrow down the possibilities by using all the relevant constraints available. Studies on word learning suggest that acquiring new words involves the use of syntax, phonology, social skills, and pragmatic reasoning (see Naigles and Swensen 2007). There is substantial evidence that children use syntax to narrow down the possible meanings of newly acquired words. This strategy, called "syntactic bootstrapping" (see Fisher, Hall, Raskowitz and Gleitman 1994; Gleitman 1990), works together with other constraints (e.g., pragmatic reasoning and visual-spatial limitations) to further narrow down the possibilities (see Markman and Jaswal 2004; Waxman 2004). There is no doubt that these learning strategies are dynamic. Children use syntax to acquire new words, and having a bigger lexicon helps augment their syntax.

Syntactic bootstrapping may be considered a learning strategy specific to the linguistic domain, since it works by establishing a correspondence between semantics and syntax. Still, word learning involves a lot more. There

is evidence that the early vocabulary is determined not only by input, but also by social knowledge, pragmatic reasoning, and phonological limitations of the child. It has been shown, for example, that early vocabulary is determined in part by the child's own productive phonology (see Fergusson and Farwell 1975), and that a child is more likely to produce a novel word if its phonology is consistent with the child's phonology than if it is not (see Leonard, Schwartz, Morris and Chapman 1981; Schwartz and Leonard 1982).

Studies on early word learning have shown moreover that language learning also follows pragmatic constraints that are relative to the speaker's intentions. From early on children focus on what the speaker has in mind, is sensitive to the communicative context, the speaker's behavior, and even to the speaker's knowledge and dispositions (see Diesendruck, Hall, and Graham 2006; Diesendruck and Markson 2001). Currently, the most promising accounts insist that language learning does not result from especially dedicated linguistic mechanisms, but from multiple mechanisms used by children for several ordinary tasks. Bloom (2000), for example, argues that word learning involves an understanding of mental states, general conceptual knowledge, and syntax.

Further studies have shown that such open interaction among domain general mechanisms is not unique to word learning, as analogous evidence can be found for speech perception and the acquisition of syntax (see Saffran and Thiessen 2007). Speech perception, for instance, exhibits categorical perception. There is categorical perception of speech when subjects are perceptually sensitive to differences across speech categories but not within categories (see Eimas, Siqueland, Jusczyk, and Vigorito 1971), suggesting that they are tuned to identify speech-relevant sounds. Yet further studies show that categorical perception is not specific for language learning, as it also appears to be used in face perception (Pollak and Kistler 2002), and it is not even unique to humans (Wang and Kadia 2001).

By the second year of age a child is already capable of using grammatical categories to learn and interact socially. There is evidence that these categories may be learned by means of domain general mechanisms of statistical analysis allowing the child to track patterns of co-ocurrence (see Mintz, Newport, and Bever 2002; Redington, Chater and Finch 1998). There is also evidence that speakers use "predictive dependencies" (e.g., a use of "the" in English predicts the presence of a noun) to identify the structure of a phrase. Predictive dependencies can be studied by using distributional information, both for linguistic and non-linguistic information (see Saffran 2001, and 2002). The analysis of distributional information is a domain general mechanism that may have shaped the structure of human languages, given that languages that have predictive dependencies are easier to learn (see Saffran 2002 and 2003).

There is also evidence against the idea that human infants passively await for linguistic input to trigger their grammatical knowledge. The evidence suggests, rather, that the child is actively participating in the process of language acquisition by using all available knowledge to interact with others. Social interaction becomes, for the active child, the pathway to language. Shatz (1994) shows how children do this by means of "elicit, entry and expansion" operations. Children elicit simple speech in transparent contexts that are rich in syntax, pragmatic, and semantic information. They are then able to store strings of speech in memory for off-line comparison. And, finally, they evaluate what they have learned by using it in further interactions.

So-called socio-pragmatic accounts (see Tomasello 2003) also support the view of an active child, attentive to language use by adults and frequently engaging in communicative interactions. Even knowledge of grammar can be accounted for in terms of domain general learning mechanisms applied to frequent communicative interactions (see Saffran, Aslin, and Newport 1996; and Gentner, Holyoak, and Kokinov 2001) together with a universal tendency to regularize repetitions (see Langacker 1987; Van Valin 1993). Grammatical competence, on this view, has more to do with having "an inventory of linguistic constructions" than with "abstract knowledge of syntax." (Baldwin and Meyer 2007, 98)

Language acquisition, the evidence suggests, results from the interaction of multiple components and different areas of knowledge. It seems best to understand language itself as a cognitive ability or mechanism that, first, results from the interaction of multiple learning mechanisms and constraints and, second, is subject to change through development. A widely accepted view within developmental psychology, owed to Shatz (1994), considers language as an interactive, constantly evolving, learning mechanism. It is through its multiple developmental changes that language becomes a powerful higher-order learning tool, granting human infants with the necessary socio-linguistic knowledge to become a person (see Shatz 1994 2007b and 2008).

Briefly put, the open view may be said to be *cognitive* in two distinct senses. Like the Chomskyan approach, it is cognitive because it focuses on knowledge of language in order to determine the very nature of its object of study, language. Now, by defending its main tenets—above described—the open view delivers an account according to which human languages are, first and foremost, cognitive and interactive tools for learning. This is a sense of being *cognitive* that is unique to this view.

This should be enough for an initial approach into the open view of language. In chapters 3 and 4, I will offer a more detailed account as I present a novel challenge to the traditional closed view (chapter 3) and present an alternative approach to language as a communicative system (chapter 4), which is

based on the notion of open compositionality. For now it will be important to recall the central tenets of the open view concerning language acquisition and development, and the two ways in which it is *cognitive*. Human languages are the product of the interaction of multiple mechanisms and domains of knowledge, and they exhibit change through development (i.e., an infant's language is not the same thing as an adult's). The view focuses on knowledge of language as the target phenomenon, and concludes that language is a powerful cognitive tool.

1.3 WHY BOTH VIEWS MATTER

I have described two distinct and widely accepted views of natural language. From within the philosophical orthodoxy the closed view understands it as a communicative system that is closed under compositionality. From within cognitive psychology and psycholinguistics the open view takes it to be a powerful cognitive tool that is both a consequence and an enabler (Shatz 2008) of the interaction among multiple mechanisms and domains of knowledge. These views seem to be orthogonal to each other, but they are not. In fact, they are intimately related.

Independently of what the closed view claims, any satisfactory account of human linguistic communication must be supported by an account of the knowledge of language that humans have. Otherwise the theory would run the risk of accounting for a language that is simply *not* a human one. The closed view is not an exception to this requirement. It does presuppose an account of linguistic cognition (i.e., acquisition and development) that is meant to support its compositionally closed account of communication. It is a well-known fact that both Frege (see Frege 1917) and Montague (see Montague 1974), two of the central figures of the view, explicitly rejected the relevance of any cognitive or psychological studies, as they considered the study of natural language (more Montague than Frege, perhaps) to be a part of logic and mathematics. Yet, more contemporary developments of the view have followed Chomsky (1986) in admitting that only by focusing on knowledge of language can we have a scientifically satisfactory account of human linguistic communication.

The closed view is nowadays the product of Fregean semantics together with a Chomskyan approach to language. The former offers the formal-compositional structure of the view, while the latter offers a framework within which the former may find proper empirical support. Philosophers of language have followed what is known as the cognitive *approach* to natural language semantics (see Chomsky 1986; Larson and Segal 1995, and Heim and Kratzer 1998, for a detailed overview; see King 2017, for recent

discussion). According to the latter, what one must focus on when studying language is the knowledge of language that speakers have—i.e., what do they know, how do they come to know it, and how do they use such knowledge. This approach is clearly in line with that of the open view. Both assume that a subject *S*'s knowledge of language £ determines the nature of £. In so doing both views deliver, or at least presuppose, an account of the acquisition and development of language.

As I said in the previous section, the Chomskyan approach is based on a particular account of language acquisition and development that differs from that of the open view. According to the former, human beings are endowed with a universal grammar that must be triggered by the appropriate environmental input (i.e., linguistic input) and can output grammatical sentences. Language acquisition is instantaneous and its development is merely a process of maturation of the universal grammar. This approach to language is itself supported by the *poverty of stimulus* and *no negative evidence* arguments (see Larson and Segal 1995, 16–19). According to the former, children learning a language £ are simply not exposed to language use that explicitly presents the underlying grammar of all possible complex expressions of £ (see Chomsky 1980). Yet, they manage to output grammatically correct expressions of £, thus suggesting that children are already endowed with grammatical knowledge of £ (i.e., a universal grammar). According to the *no negative evidence* argument, when children learning a language £ make mistakes, they are typically not corrected and when corrected they pay no attention to corrections, thus suggesting that children do not recover from mistakes by learning from others. Once again, it seems children are already endowed with all that is needed to be competent speakers of £.

The claim that language is closed under compositionality and is, hence, isolated from external—i.e., non-linguistic—processing/cognition fits well with the claim that language is a separate cognitive mechanism, perhaps genetically endowed, in need of little more than triggering by the proper linguistic input and with little to no substantial change owed merely to maturation. If language is such an independent and autonomous cognitive mechanism, then compositional closure appears to be a natural syntax/semantics consequence. After all, on this view knowledge of language is itself domain specific, separate, and autonomous. It is in need of no aid from other learning mechanisms or areas of knowledge and, thus, may be isolated from any external—i.e., domain general—influences.

Yet, as we saw in the previous section, multiple empirical studies have shown that the arguments for the Chomskyan proposal are not as sound as they seem to be. It has been shown that the learning child is not a passive input receiver, but an active learner (see Shatz 1987) that can draw important insights even from a relatively small set of data (see Elman 2003). Other

studies on infant-directed speech have shown that such input does in fact provide substantial clues, both syntactic and semantic (see Jusczyk 1997; and Kuhl 2004). These findings have given place to a debate about the abstract nature of the underlying grammar and its relation to input, with proponents of the Chomskyan view claiming that input merely triggers the setting of certain parameters already included in a highly abstract universal grammar (i.e., too abstract to account for semantics) and critics defending that grammar is constructed (at least partly) from the input (see Mueller Gathercole and Hoff 2007, for a detailed review). Either view, however, must account for the evidence showing that language learning involves a complex and rich input and a highly active learner—i.e., evidence that contradicts both the *poverty of stimulus* and the *no-negative evidence* arguments.

More importantly, the evidence strongly suggests (see section 1.3) that language is not an independent, domain specific cognitive mechanism; that its components are not autonomous; and that knowledge of language is acquired through the interaction between multiple distinct learning mechanisms and domains of knowledge—e.g., statistical learning, distributional information analysis, structure mapping, pattern recognition, and understanding of mental states, among others. Given that language as a communicative system is the product of such complex and highly interactive process of acquisition and development there is little reason to expect such system to be autonomous, independent, and, even less, closed to any so-called external processing or information. To begin with, there is no specific "language mechanism" or "language apparatus" in charge of attaining and developing knowledge of language. Hence, there is no domain specific mechanism in relation to which domain general reasoning (e.g., pragmatic reasoning, statistical learning, or intention understanding) may be said to be external. Thus, the idea that language use is closed under compositionality, that it involves the use of linguistic and only linguistic processing, is left wanting in empirical support. In chapter 3, I will argue that the empirical evidence does in fact pose a substantial challenge against the closed view.

I have said that the closed view requires an account of how humans come to have knowledge of language and, hence, an account of language acquisition and development. This is already enough to see how relevant the open view is for the closed one. I believe, however, that there are further reasons to think that both views are relevant for each other and, furthermore, that they are intimately related. It seems to me that each view is based on central claims that explicitly appeal to the expertise of the other. Consider first the closed view. It is meant to be an account of how humans manage to communicate by using a productive and systematic representational system. According to this account, humans manage to do this in virtue of *knowing* a compositionally closed language constituted by a finite lexicon, syntax, and

semantics—including syntax-semantics recursive rules for complex expressions. Humans, with their finite cognitive resources, manage to become linguistically *competent* in virtue of *knowing* a finite yet infinitely productive system. Briefly put, like the open view, the closed view is itself a theory of what a competent speaker knows in virtue of her linguistic competence.

Consider now the open view. It is meant to be an account of what humans know when they know a language. More specifically, it is a detailed account of what human infants learn, and how they learn it, in order to become adult language users. According to this account, human infants do so by using all available knowledge, from varied areas and domains, together with multiple learning mechanisms. One crucial element of such learning process is the use of socio-linguistic knowledge to participate in linguistic communication. The cognitive view claims that humans eventually become competent by participating in communicative endeavors. Thus, the open view delivers a theory of what speakers need to know to engage in linguistic communication, and so does the closed view.

Of course, there are important differences between the views. The closed view focuses exclusively on the knowledge needed for communication and says nothing about its acquisition and/or development; it merely assumes that such knowledge is somehow attained. The open view strives to explain how is such knowledge possible and says little about the nature of linguistic communication; it merely assumes that such interaction somehow takes place. It is because of this that the open and closed views are complementary. Our current understanding of language, both knowledge and communication, can only benefit from interdisciplinary work with both views.

The open view presents us with what is nowadays our best account of human knowledge of language, in terms of language acquisition and development. Given its wide empirical support, it constitutes our best approach to understanding natural language. Similarly, the closed view constitutes our most satisfactory account of linguistic practice. It does not only explain how speakers manage to convey their thoughts to others and understand the thoughts conveyed by others, it is also the only account of how natural languages manage to be *productive*—so that humans, with finite cognitive abilities, can understand an unlimited number of complex expressions—and *systematic*—so that speakers that understand the expression "A loves B" thereby understand the expression "B loves A." Linguistic communication, productivity and systematicity are unavoidable *explananda*, even for the open view of language. The closed view should strive to show how the empirical evidence on what actual humans know when they acquire a language supports its compositional understanding of communication. The open view should strive to show how a satisfactory account of linguistic communication, productivity, and systematicity could result from its interactive, multiple

domains account of language acquisition and development. Both goals can be achieved by a single effort to present a compositional account of communication within an open view of knowledge of language. This, in a few words, is the idea of open compositionality, which I will develop in chapter 4.

1.4 TOWARD OPEN COMPOSITIONALITY

In this book I argue that we can substantially improve the closed view of language by abandoning the commitment to compositional *closure*—i.e., the idea that the content of any complex expression is *fully* determined by the meaning of its parts and the way they are combined—while endorsing an *open* view towards compositionality—i.e., the idea that compositional processes typically interact with multiple non-linguistic, domain general, mechanisms and reasoning strategies to determine the content of a complex expression. I call this view "open compositionality." According to this view, compositional processes are only one among multiple different procedures required to account for the meaning of complex expressions. On this proposal, compositional processes involving the operation of grammatical (i.e., syntax-semantics) rules are open to interact with other cognitive processes—e.g., Theory of Mind Mechanism, socio-linguistic knowledge, pretense abilities, pragmatic reasoning, and different language learning strategies, among others. This is where the proposal departs from the tradition. On this view compositionality is *not* always *enough* to account for the meaning of a complex expression. Furthermore, it is sometimes not even necessary.

Open compositionality claims that compositional linguistic processes interact with various non-linguistic, domain general processes, and that one must look at cognitive and developmental psychology to determine which are those processes and how they interact. To understand this view it will be important to distinguish it from other proposals that may appear to be similar. It is important to distinguish open compositionality from *interactive* compositionality. The former takes compositionality to be necessary and interaction to be possible—i.e., compositional processes *are open* to interaction. The latter takes it to be necessary. Open compositionality is compatible with there being cases where purely compositional processes are enough to determine content as well as cases where it plays no role; interactive compositionality is not.

Open compositionality differs from non-compositionality. This is a crucial distinction that is very commonly ignored. To say that language is compositionally open is *not* to say that language is *not compositional*. Non-compositional accounts are typically associated to semantic holism or, more recently, conceptual-role semantics (see Quine 1960; Wittgenstein 1967; Brandom

1994; and Sellars 2007). According to such views the meaning of a complex expression of £ is determined by the totality of meaningful expressions in £ and how they relate to its constituent expressions. According to such view, individual expressions do not have a meaning of their own, or independently from other expressions—or even independently from human experience. It follows that purely linguistic compositional processes *will never* be enough to determine the meaning of a complex expression. Contrary to such non-compositional accounts, open compositionality does claim that individual expressions have a meaning of their own, and that knowing such meanings may sometimes be enough (together with the relevant compositional processes) to determine the meaning of a complex expression. What open compositionality rejects from the orthodox view is the associated idea of *closure*; it rejects the idea that a complete compositional story is all there is to be said about meaning determination for any complex expression. Unlike semantic holism, open compositionality *does not* reject the idea of compositionality itself.

Open compositionality also differs from what is nowadays known as "truth-conditional pragmatics" (see Recanati 2010 and 2012). According to truth-conditional pragmatics, purely linguistic compositional processes *necessarily* interact with pragmatic reasoning—i.e., reasoning about the intensions of speakers—to determine content. On this view, compositional processes can never fully determine content. "To get full-blown truth-conditional content, pragmatics will be needed." (Recanati 2010, 3) Open compositionality differs from this latter view in three important respects. It does not consider the interaction to be *universal*. The non-linguistic processes that may be relevant for content determination include much more than just reasoning about the intentions of others. And it allows for cases in which the relevant content is not compositionally determined *even after non-linguistic processes takes place*. In other words, open compositionality is compatible with cases for which traditional compositional processes play no role. Truth-conditional pragmatics, however, claims not only that a semantics-pragmatics interaction is necessary, it also claims it is *sufficient* to determine the content of all complex expressions. Thus, as Recanati (2012) points out (see also Pagin 2005; and Westerståhl 2012) truth-conditional pragmatics can be seen as offering a *compositionally closed* account, albeit one that includes pragmatics, together with syntax and semantics, as part of the compositional processes. Assuming that modulated meaning is the product of semantics-pragmatics interaction, "the modulated meaning of the complex (. . .) is a function of the modulated meaning of the parts (and the way they are put together) *plus*, *in addition, the context which determines how the content of the whole itself is modulated*." (Recanati 2012, 190) As I will show in chapters 3 and 4, there are cases requiring more than just pragmatic reasoning—in

which semantics-pragmatics interaction is *not* sufficient—and cases in which there is no role to be played by compositional processes.

I believe open compositionality offers a more satisfactory account of the content for simple and complex expressions of natural language than the orthodox compositionally closed account. The empirical evidence better supports it and it offers explanations of linguistic phenomena where the orthodoxy has only found trouble. These substantial theoretical advantages are owed to a methodological shift. As I will show in chapter 2, the orthodox view starts by assuming that natural languages are not so distinct from formal languages and, in particular, that they are compositionally closed. This assumption works at a methodological level, helping theorists identify the needed syntactic / semantic structure for the relevant expressions to be handled by a compositionally closed language. As a methodological assumption, the hypothesis of compositionality tells us that whenever a given expression appears to behave non-compositionally—e.g., when a complex expression involving proper names (Frege 1892) or modals (Kratzer 2012) delivers different meanings in different contexts—we must look out for non-obvious syntactic or semantic complexity. Following this methodology closed compositionality becomes a goal, not an empirical discovery about natural languages, for it is known that the content of any complex expression may be given a closed compositional account, provided one is free to postulate the needed syntactic /semantic complexity among the constituent expressions (see Zimmermann 2012; and Pagin 2012). As I will argue in chapter 3, this methodology gives place to *ad hoc* postulations lacking independent empirical support, eventually making it improbable to build a theory that is compatible with what we know about language acquisition and development (i.e., actual human knowledge of language).

Unlike the traditional view, open compositionality starts by looking at empirical studies on language acquisition and development, as well as on cognitive development, to determine what speakers actually know in virtue of being competent speakers of a given natural language. Only after this is achieved, claims open compositionality, can we go on to offer compositionally closed or openly interactive accounts of the relevant expressions. The hypothesis of compositionality may be useful and fruitful, but it must be subject to empirical revision, like any other hypothesis that aims to explain empirical phenomena such as human linguistic communication.

An important consequence of this methodological shift is a substantial increase in theoretical tools. Not only are there advances in logic and syntax that can be of great help to improve our semantics, we also have theories of word and syntax acquisition, theories of pretense and the understanding of mental states, theories of social cognition, theories of memory and its linguistic application, and much more that proves to be relevant for determining a

speaker's knowledge of language and how she comes to know it. Because of this methodological shift and its theoretical consequences, open compositionality stands on a wider, more varied, and more complex empirical basis than that of its closed ancestor. Furthermore, in virtue of its openness to include non-compositional processes to account for the determination of content for complex expressions, open compositionality is not forced to postulate *ad hoc* semantic or syntactic complexity. Whether and to what degree natural language is a compositional system is a matter of empirical discovery, not a pre-established theoretical goal, on this view. As I will show in chapter 4, open compositionality is in a better methodological and empirical standing than the orthodox *closed* account.

Open compositionality presents a more complete, inclusive, and satisfactory understanding of natural language, or so I will argue in this book. In chapter 2, I will present the traditional view of language as a compositionally closed system of communication. I will present its central theses and methodology, and show how they are applied to account for different problematic phenomena. I will then identify the cognitive commitments of the view in order to get a clear picture of what language is as a human cognitive capacity—how it is processed, understood, acquired, developed, etc.—according to the closed view (see Baggio, van Lambalgen, and Hagoort 2012). This will help the reader get a clear and more complete understanding of the traditional view of natural language as compositionally closed. The chapter concludes by presenting the three *desiderata* most commonly wielded on behalf of the orthodox view. These *desiderata* constitute explanatory challenges that any satisfactory account of language should meet.

In chapter 3, I will test the empirical standing of the closed view by operationalizing the principle of compositionality, thereby showing that it has two substantial empirical commitments, a cognitive and a processing one. I will then consider evidence on language acquisition and development, as well as studies on language processing in adults, to show that the empirical evidence falsifies both commitments. It turns out that *closed compositionality* fails both as an account of what competent speakers know when they know a language, and as an account of what speakers do when they understand and process a given natural language.

In chapter 4, I present an alternative view of natural language, *open compositionality*, accompanied by a novel non-compositional *cognition-first* methodology. This account observes what I call the *Lewisian Compromise* according to which "Natural languages are, first and foremost, things that can be learned, developed and used by human beings given their limited cognitive resources." *Open compositionality* is developed by explaining what we know when we know a language and how we know it—i.e., a supermodular cognitive capacity by means of multimodular interaction—and how we use

this knowledge—i.e., sentence interpretation as a decision-making task. *Cognition first* offers a methodology, an alternative to traditional compositional ways of solving theoretical problems of natural language, that guides us to first understand the underlying cognitive processes, as understood by our best cognitive psychological and psycholinguistic theories (among other cognitive scientific efforts) before attempting to explain any problematic linguistic feature. I show how this novel approach to language can successfully account for the productivity, systematicity, and computability of natural languages, and conclude by contrasting *open compositionality* against other alternatives available in the literature that also oppose *closed compositionality*.

The remaining three chapters are intended to display the explanatory power of *open compositionality* and the *cognition first* methodology by showing how their successful application results in satisfactory accounts of otherwise problematic phenomena. Chapter 5 offers an account of the substitution failure phenomena—i.e., whenever the substitution of correferential names fails to preserve the meaning of a given complex expression. I look into studies on memory for proper names, heuristic strategies for acquiring new proper names and, most importantly, studies on the processing architecture for storage and retrieval of proper names. The evidence suggests a processing architecture account of why competent speakers may process proper names differentially, regardless of whether the said names differ or not in denotation, thus explaining why and how substitution failure takes place.

Chapter 6 offers a cognitive theory of how empty names—i.e., names that are meaningful even though they have no denotation—are competently used to convey non-trivial information. I look into empirical studies on pretense, most importantly Leslie's seminal contribution (see Leslie 1987), to see how speakers may benefit from the cognitive apparatus needed for pretense—i.e., a representation decoupler—by applying it to communicative scenarios involving linguistic expressions. This delivers an account according to which empty names are given surrogate interpretations that make them meaningful without thereby assigning a genuine denotation. This chapter was previously published as "A cognitive theory of empty names" in the *Review of Philosophy and Psychology*. I am thankful to the editors at Springer for granting the rights to reprint it here.

Finally, in chapter 7, I go beyond the purely linguistic domain to show how the *cognition first* methodology may be greatly useful in other areas of philosophical interest. This chapter is concerned with the problem of moral knowledge and associated metaphysical, epistemological, semantic, psychological, and mental complications. I describe the problems posed by moral knowledge and then offer a detailed, although inevitably partial, review of the available studies on moral cognition, moral development, and associated

evolutionary studies. These studies strongly suggest that moral knowledge is fundamentally determined by the nature of human cognition. I argue, furthermore, that these studies show that morality and language are strongly analogous cognitive abilities and should be understood as such. I present the profound implications of this analogy and suggest that moral knowledge is better understood in relation to moral competence. I describe how moral competence may be said to constitute and/or give place to moral knowledge. I offer a description of the relevant (psychological) moral facts and show how this is compatible with the universality and objectivity of some moral principles. The result is a broad picture of a yet to be developed naturalistic moral realism that is based on the idea that what is special about morality is moral cognition, not the existence of peculiar, mind-external, moral facts.

NOTE

1. I believe some such view can benefit both, philosophy of language and cognitive psychology/psycholinguistics, yet throughout the book I will focus on the philosophical side for two reasons. First, because serious consideration on how to implement the lessons from philosophy to improve our understanding of language acquisition and development requires careful experimental planning and design, none of which is within the scope of this book. Second, because some of it has already been done, with cognitive scientists using philosophical insights to guide their research (e.g., Semenza 2009).

Chapter 2

The Closed View and Strong Compositionality

To properly understand why students of language should move from a *closed* to an *open* yet compositional view we must first get a proper grasp of the view we are departing from. The purpose of the present chapter is to present and develop the closed view with as much detail as needed. I will begin (section 2.1) with a brief historical review that will help us understand the origins and motivations behind the contemporary sway of the principle of compositionality. Section 2.2 distinguishes among different versions of the principle and shows how it is meant to work methodologically speaking. Section 2.3 illustrates the principle *qua* method by looking into a well-known debate in philosophy of language—i.e., context sensitive expressions. Section 2.4 considers several positive arguments on behalf of compositionality that are meant to offer independent empirical support for the view. The chapter concludes (section 2.5) by pointing out the cognitive assumptions of the closed view, suggesting that they might not be compatible with what we know about human language acquisition (see chapter 1, sections 1.2 and 1.3). This will set the stage for chapter 3, where I will present a novel objection against the closed view, based on evidence from language acquisition, language development, and the historical development of human languages.

2.1 A (VERY) BRIEF HISTORY OF COMPOSITIONALITY

Let me begin with a brief historical overview of compositionality. Doing so will help us better understand the closed view of natural language and, also importantly, how we got here. According to Janssen (2001), discussions about compositionality date at least as far back as the early nineteenth century (though see Szabó 2013; according to him it goes as far back as Leibniz

1646-1716). Towards the end of the eighteenth century, the most widely held view was rather opposed to the ideas associated to the principle of compositionality. Kant (1724-1804) and Schleiermacher (1768-1834) argued for a primacy of the complex expression (i.e., the judgment in Kant's case) over the expression parts. On this view the latter somehow inherit their meaning from the former. This view is commonly associated to an apparently opposing principle to that of compositionality, this is commonly known as the principle of *contextuality*.

Contextuality: The meaning of a simple expressions *e* of a language £, can only be determined within the context of some complex expression $<E_1, E_2, \ldots E_n>$ of £ in which it may appear.

This principle appears to contradict the principle of compositionality, according to which the meaning of a complex expression is fully determined by the meaning of its parts and the way they are combined. Yet, there is controversy as to how contextuality and compositionality relate (see Janssen 2001 and 2012) to each other. What seems clear (see Janssen 2012) is that both principles were well known to mid-nineteenth century German philosophy, as they both are discussed in influential texts, such as Trendelenburg's *Logische Untersuchungen* (1840). Half a century later, this debate reached Frege, by means of Lotze (1874) and Wundt (1880), whose foundational *Über Sinn und Bedeutung* (Frege 1892) is widely—and according to Janssen 2001, mistakenly—viewed as championing the principle of compositionality. Like other nineteenth century German logicians, Frege appears to endorse both the principle of compositionality *and* the principle of contextuality, which he explicitly defends and applies in his *Foundations of Arithmetic* (1884). The principle of compositionality would not become the orthodoxy until mid-twentieth century, and understanding the causes of this historical change gives us an important insight into the nature and limits of the, now orthodox, *closed* view of language.

The extraordinary shift from the primacy of contextuality in the early nineteenth century to the rule of compositionality in contemporary philosophy of language was made possible by the creation, development, and continuous improvement of mathematical logic, from the late nineteenth century onwards. Frege (1879) pioneered such historical change by producing one of the first works on symbolic logic in the tradition. With the help of symbolic logic the principle of compositionality proved to be more useful than its rival. Decades later, Church (1941)—among others—helped formalize the study of natural language, while Carnap's (1947) seminal intension / extension distinction offered a new formal methodology (i.e., model theoretic semantics) with which to approach all of natural language. Carnap (1947) is perhaps the first logician that explicitly endorses the principle of compositionality

as central to the study of language. He also explicitly rejected contextuality, "A decisive difference between our method and Frege's consists in the fact that our concepts, in distinction to Frege's, are independent of the context." (Carnap 1947, 125)

This mathematical and formal approach to language would be further refined and more successfully applied after Carnap's (1947) work on intensional logic. Not surprisingly, the goal of the study of language, on this mathematical approach, is set in logical and mathematical terms. The opening statement of Montague (1970b) is quite clear on this.

> There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed, I consider it possible to comprehend the syntax and semantics of both kinds of languages within a single natural and mathematically precise theory. (Montague 1970b, 222)

What does it mean to "comprehend" the syntax and semantics of a natural language "within a mathematical theory"? The answer is simple as well as surprising. For Montague, as well as for most philosophers of language endorsing the closed view, one has an account of the syntax and semantics of a natural language if and only if one has a mechanical, abstract way to build all the syntactically and semantically possible constructions of that language. The job of a theory of natural language is, on this view, that of providing the set of rules that determine how the syntactic and semantic symbols are manipulated to construct complex expressions. This set of rules would thereby comprehend the syntax and semantics, in virtue of being an algebra of the relevant language.

It is expected, then, for this tradition to consider natural languages as abstract entities. As I said in chapter 1, when presenting the closed view, languages are taken to be sets of ordered pairs of sentences and meanings. The goal of a philosophical theory of language, from this perspective, is to account for the function that delivers the correct set of such ordered pairs for a given language. If we have the proper algebra, we should be able to offer a mechanical way to do this (see Lewis 1975). That is precisely what Montague (1974) achieves, thus explaining why his work has been so influential among philosophers and logicians and, thanks to the successful use of Montague's mathematical approach by Partee (see, for example, Partee 1984), also among linguists.

Montague's (1974) proposal assumes that each simple expression can be given a meaning in isolation (i.e., independently of its being used in any complex expression), and that there is a one-one correspondence (more specifically a homomorphism) between syntactic and semantic rules. Doing so would help reach our algebraic goal by achieving a more specific, sentence-related goal.

> The basic aim of semantics is to characterize the notions of a true sentence (under a given interpretation) and of entailment, while that of syntax is to characterize the various syntactical categories, especially the set of declarative sentences. It is to be expected, then, that the aim of syntax could be realized in many different ways, only some of which would provide a suitable basis for semantics. (. . .) I fail to see any great interest in syntax except as a preliminary to semantics. (Montague 1970c, 223 fn)

To sum up, to account for a natural language £ is to offer a mechanical and mathematically elegant way to determine the truth conditions for any possible sentence of £ in a way that observes the proper logical relations—e.g., entailment—among sentences. It is thus not a surprise that, on this view, syntax is nothing more than a preliminary to semantics, since it is only the latter that determines truth conditions. As proposed by Carnap (1947) and Montague (1974), this account has been further applied and developed by Lewis (1975) and (1980), Kaplan (1989), Kamp (1981), Kamp and Reyle (1993), Partee (1975), Kratzer (2012), and so many others after them.[1]

It should be clear why any language £ that is studied in this way *must* be seen as compositionally closed, for otherwise the algebraic goal becomes unattainable. If £ is not compositionally closed then it must be that the meaning (or truth conditions) of a complex expression E, or class of complex expressions $<E_1, E_2, \ldots E_n>$, of £ is partly determined by factors that are, first, neither syntactic nor semantic, and second, cannot be predetermined by the syntax or the semantics—i.e., the algebra of £. This could happen, for example, if the meaning of E is, at least sometimes, determined by nonlinguistic (neither syntactic nor semantic) cognitive general resources, such as pragmatic or pretense-like reasoning. But not only, it can also happen if the given syntactic structure or semantic value of E is itself a product of such cognitive-general reasoning, as opposed to being determined by the syntactic and semantic algebra of £. Either way, there would be no mechanical way to determine the meaning of the relevant expression E or class of expressions $<E_1, E_2, \ldots E_n>$ of £ and, hence, it would be impossible to offer a mechanical way to construct *all* possible sentences of £. Those involving E, or any member of $<E_1, E_2, \ldots E_n>$, would be left unaccounted for.

The latter should help illustrate how important the *closed* shape of natural language is for this view. Anything distinct from full compositionality for £ is tantamount to rejecting the possibility of offering a *mathematical* account of £. And this, for the closed view of language is equivalent to rejecting the possibility of *any* compositional account of £, full stop. If this were true, if the rejection of the mathematical view meant the impossibility of any compositional account, the consequences of not having *closed compositionality* would be truly undesirable. If no *compositional* account were possible, we could not even explain how a human being can acquire a language. If meaning is

not at all compositionally determined, competent speakers must somehow get to know the meaning of complex expressions (or of the totality of such expressions) before they can know the meaning of words. Given that human languages have infinitely many complex expressions, the task of acquiring a language appears to be impossible. In fact, the most common—perhaps also the best—argument for the closed view (see Szabó 2012) claims that only by assuming *closed compositionality* can we explain the productivity, systematicity, and simplicity of natural language. I will be considering the validity of this argument in section 2.5.

I believe, however, that it is a mistake to equate a *mathematical* account of language with a *compositional* one; not all compositional accounts are mathematical. As we have seen, the latter demands that languages be compositionally *closed*, in that the meaning of all complex expressions must be determined by the syntax *plus* semantic algebra. Yet, one might deny this while still claiming that language is, in some partial way perhaps, a compositional system of representation. Even if *not all* expressions get their meanings determined by the syntax *plus* semantics rules, it might still be that complex expressions *may* get their meaning determined by their parts (and the way they are combined). Furthermore, it might be that—as I suspect—we can identify the exceptions to this generality and explain how or why they take place. This would not be an elegant *mathematical* account with a powerful mechanical way to determine the truth conditions of every possible sentence of any natural language. But it would still be, at least, a *compositionally weak* theory capable of accounting for language acquisition and development, language productivity and systematicity, and a host of other phenomena that seem to go against closed compositionality (see chapter 3). So, unless one explicitly desires an algebraic account of truth conditions for all possible sentences of English for the sake of it, one might find comfort in non-mathematical accounts that are *nonetheless* compositional (see chapter 4).

2.2 THE PRINCIPLE(S) AND METHODOLOGY

The closed view of language is in the business of offering a mathematical account of meaning for complex expressions. To do so, it must rely on the principle of compositionality. But, what does that principle say exactly? And, how exactly does the theory rely on it? I will answer these questions in this section.

There are multiple versions of the principle of compositionality. These versions vary depending on their scope—some are limited by immediate syntactic structure, others are aimed at psychological explanation—and

strength—some aim at all possible structures, others do not. Szabó (2012) distinguishes among three versions, depending on strength, and three more depending on scope. On its most intuitive version, the principle of compositionality is weak and uncontroversial. This is perhaps the most well-known version of the principle, owed to Partee.

Weak Compositionality: The meaning of a compound expression is a function of the meanings of its parts and of the way they are syntactically combined. (Partee 1984, 281)

This version has the virtue of being flexible enough to allow for different ways in which complex meanings may be composed. It merely states that complex meanings are *a function* of the meanings of expression parts. This does not exclude the possibility that they also be a function of something else, in addition to the meanings of expression parts. In this sense, it is hard to see how natural language could *not* be compositional. Even if there are further determining elements, beyond syntax and semantics, complex meanings may still be a function of the meaning of expression parts. This, however, is seen more as a vice than a virtue from the perspective of the *closed* view for a rather simple reason: weak compositionality is not enough to deliver an algebraic account of complex meanings for natural language.

Szabó (2012) points at three different problems associated to the weak version of the principle. First, as I have already mentioned, the weak version does not say what type of function relates the meaning of a complex expression with the meaning of its expression parts. This could be a partial determination relation, a complete determination relation, a constitution relation or any other. If we are looking for the set of syntactic and semantic rules for £ that, together with £'s lexicon, may offer a mechanical way to determine the meaning of any possible complex expression *E* of £, only a complete determination relation will help. This delivers a first strengthening of the principle.

First Strengthening: According to the principle of compositionality, the meaning of a complex expression *E* of £ is fully determined by the elements described in the principle.

A second problem arises when we consider what the meanings of the expression parts could be. There are at least two ways in which we could be talking about such meanings. We could be talking about the meanings that the expression parts have *in isolation*. In other words, we could be talking about the meanings that the expression parts have independently of how they relate to other expression parts of the complex expression *E* of which they are part. We could be talking about the meanings that the expression parts *collectively* have as parts of *that* complex expression, this could include both the meanings the expression parts have in isolation together with meanings that

may arise in virtue of the presence of other expression parts and how these combine with each other in order to constitute *E*. Like a partial determination relation, *collective* meanings of expression parts are unwelcome in a compositionally closed view. If the meaning of a complex expressions like *E* were to be determined by the collective meanings of their expression parts $<e_1, e_2, \ldots e_n>$, we would not have a mechanical way to determine complex meanings for a language £, for we would first have to determine collective meanings for expression parts, and these could vary depending on the complex expression *E* in which the expression parts $<e_1, e_2, \ldots e_n>$ may appear. Thus, we need the meanings of expression parts to be understood as the meanings they have *in isolation*. This delivers the *second strengthening* of the principle.

Second Strengthening: according to the principle of compositionality, the meanings of expression parts to be considered are those they have *in isolation*.

Finally, a third problem arises from an ambiguity in the expression "they" as it refers to whatever gets to be combined in some or other way according to the principle. There are two possible antecedents for the expression. First, it could be referring to the meanings of the expression parts, telling us that the meaning of the complex expression *E* is a function of the way in which the meanings of its parts are combined. This would be tantamount to claiming that complex meanings are only determined by semantic, not syntactic, compositionality, effectively creating an independence of the former from the latter. If the meaning of *E* is determined only by the way the meanings of $<e_1, e_2, \ldots e_n>$ are combined, the latter could semantically combine in ways that do not respond to the syntax of *E*, thereby assigning multiple arbitrary complex meanings for *E*. To illustrate consider the sentences E_1: "John loves Mary" and E_2: "Mary loves John." Both have exactly the same expression parts < "John," "Mary," "loves">; yet they clearly have a different meaning. But if semantic composition is to be considered independent from (or prior to) syntax, it is easy to combine the expression parts of E_1 and get the same complex meaning, (Mary loves John), that we would assign to E_2. This is clearly a negative result. The things that are combined in the relevant ways to compose the meanings of complex expressions cannot be the meanings of the expression parts independently of their syntactic arrangement.

Alternatively, the expression "they" could refer to the expression parts. On this reading, the principle would be stating that complex meanings are a function of the *syntactic* way in which the expression parts are combined. Accepting this reading would be equivalent to claiming that syntactic compositionality comes first, while semantic compositionality is dependent on it. This is perhaps the most traditional way to read the principle of compositionality. On this reading, semantic combination rules are dependent on syntactic

ones. If this were the case, then we could imagine how a mechanical account of complex meanings could go, ideally assigning first a unique syntactic way of combination and, later on, a unique semantic one, thereby excluding the possibility of assigning arbitrary interpretations. This gives place to a third strengthening of the principle, one that gives us a *vertical* and *bottom-up* direction of the compositional process.

Third Strengthening: According to the principle of compositionality, the compositional progression proceeds from syntax to semantics and from expression parts to the complex expression.

Once we have fixed the principle to make sure that it is compatible with a mathematical account of complex meanings for natural languages, the result is a rather strong principle of compositionality:

Strong Compositionality: The meaning of a compound expression *E* of £ is completely determined by the meanings the expression parts $<e_1, e_2, \ldots e_n>$ of *E* have in isolation in £, and the syntactic way the expressions are combined.

On this *strong* version, the principle of compositionality is read as claiming "that once you fix the meanings of the constituents of an expression and its syntactic structure, you have fixed what the expression means." (Szabó 2012, 71) According to this interpretation, if a language £ observes the *principle of compositionality*, then the meaning of any possible complex expression of £ depends on the lexicon of £ (i.e., the meaning of expression parts in isolation), on syntax, and, most importantly, on *nothing* else. As I said before (see section 1.1), the resulting principle is *logically* closed, given the universal scope of all its quantifiers. It is meant to apply to *all* complex expressions, and the account of how the meaning of such expressions is determined is also an account of *all* possible meanings for such expressions. As I will show in the following chapter, this logical *closure* has substantial consequences having to do with the nature of linguistic knowledge and processing.

To better understand how *strong compositionality* works on behalf of the mathematical view of natural language we should keep in mind some further assumptions. According to Janssen (Janssen, 1997, 426-427), the following six assumptions must be true of £:

- *Syntax and Semantics*: the syntax and the semantics of £ must be separate components, connected only by compositionality.
- *Output—Input*: the output of syntactic operations in £ *is* the *only* input for semantic determination in £.
- *Part individuation*: which expression *parts* a complex expression *E* has in

£ is determined by the grammar of £. The individuation of parts may not coincide with speaker intuitions.

- *Rules*: combination rules of £ govern both syntactic and semantic combination. Expression parts in £ can only be combined in the way determined by the syntactic rules, and complex meanings in £ are determined only by the meanings of expression parts in £ and by the combination rules of £.
- *Rule correspondence*: every syntactic combination rule of £ has a corresponding semantic rule describing how the former has an effect on the meaning of a complex expression in £. To achieve this correspondence we can tweak either the syntactic or the semantic rules (following either semantic or syntactic considerations) in £.
- *Meaningful parts*: no expression that is identified as a *part* may lack a meaning.

In a nutshell, the *compositionally closed* view of a language £ takes all possible complex expressions to have a corresponding meaning that is fully, and only, determined by the *grammar* of £. The grammar of £ is, in turn, constituted by the lexicon (or vocabulary) of £, the syntactic rules of combination, and the semantic rules of combination. In this sense of grammar, the grammar of £ is both necessary and sufficient to determine the meaning of any possible complex expression of £. Nothing else—i.e., beyond lexicon, syntax, and semantics—is needed; and nothing else may achieve this goal.

This delivers a very clear and simple view of language that is accompanied by an equally simple methodology for theorists to follow as they struggle to identify the grammar of a given natural language. Suppose, for instance, that you are trying to account for the grammar of English, as Montague does for a fragment of it (see Montague 1970a; 1970b; and 1970c). As you move on you identify what appear to be problematic complex expressions of £ that do not appear to observe the principle of compositionality, either because they seem to vary their meaning in relation to non-grammatical influences, or because they appear to have meaningless expression parts, or for some other reason. What should you do? The answer is rather straightforward. Since, by assumption, £ is compositionally closed, its syntax and semantics must be kept apart, though with corresponding rules of combination making the output of the former become the input of the latter. The grammar that you are aiming at works by identifying which elements count as an expression part; making sure that all such expression parts have an assigned meaning; fixing the relevant syntactic and semantic rules; and applying these rules. This formal proceeding gives you, the theorist, at least three different dimensions of flexibility among which you may find enough leeway to change or modify your grammar to make sure that £ remains compositionally closed. First, you may individuate expression parts differently, e.g., you may consider strings of words to be a

single expression. Second, you may postulate as many meanings as necessary, e.g., you may consider fictional names to have fictional referents. Third, you may individuate the syntax (or semantics) of a given expression in whatever way is needed to make it correspond with the semantics (or syntax) of that expression; e.g., you may consider that some simple expressions do have a complex syntactic structure. Which of these problem-solving strategies you, the theorist, should follow may vary from case to case, and sometimes you may find it convenient to follow more than one (see Zimmermann 2012, for a detailed account of these strategies).

We now have a better idea of what the principle of compositionality claims, according to the closed view of language, and how it is meant to constrain our theorizing. In the following section I will present some well-known examples of theorists (philosophers, linguists, and logicians) applying this methodology. It should be clear that the resulting principle could easily avoid the problems noticed by Szabó (2012), yet it does so at an important cost. Unlike its weak sibling, *strong compositionality* has little intuitive import, and it is so specific (only lexicon and syntax/semantics) and at the same time universal (all possible meanings for all possible complex expressions) that it is difficult to find direct empirical evidence supporting it. Furthermore, and this is perhaps the most problematic aspect, there is so much room for maneuver to accommodate problematic phenomena that it is almost impossible *not* to deliver a compositional account. In fact, it is possible to prove that, if you are willing to accept an unnatural grammar (Janssen 1997), a set of complex enough meanings (Zadrozny 1994), or certain restrictive conditions granting polynomial compositionality (Pagin 2012), "any language may be given a compositional semantics." (Janssen 2012, 42)

Mathematical results proving that any language £ may be given a compositional semantics, provided one is willing to accept an unnatural grammar, may be seen as evidence that the principle of compositionality is vacuously true. This and other problematic aspects of the principle of compositionality have given place to further debate concerning its nature and theoretical use, giving place to three roughly distinct positions. First, there is the methodological stance. Some (see Janssen 1997; and Jacobson 2014) have argued that the principle should be seen as a methodological one, guiding theorists in their quest for a semantic account of a given language. On this view, "the challenge of semantics is to design a function that assigns meanings, and (. . .) the best method is to do so in a compositional way." (Janssen 1997, 457) Second, there is also an empirical take on the principle. Theorists in this group (see Szabó 2012; 2013; Dowty 2007) have argued that the principle should be considered as having an empirical import and that the relevant grammar should observe certain empirical limitations, making it more natural and the principle less vacuous. Supporters of this empirical stance typically

offer three arguments intended to show that natural languages, as a matter of empirical fact, do observe the principle of compositionality. These arguments are based on the alleged *productivity*, *systematicity*, and computational *simplicity* of natural languages. I will offer a detailed account of these arguments, as well as the methodological one, in section 2.4.

The third, and last, standpoint on the principle (see Partee 1982, and 1988) considers it a methodologically important principle that, nonetheless, cannot be seen as offering an account of human languages. The principle of compositionality, and its accompanying methodology, abstracts away from many important features of natural languages, endowing them with a more rigid, closed, and fixed semantics than they really have. This standpoint is a much more amicable one to the view that I want to defend. Yet, I fully reject the principle and propose an alternative methodology in chapter four, where I present the view I call "open compositionality."

2.3 APPLYING THE METHODOLOGY

Given its endorsement of *strong compositionality* and its accompanying assumptions, the closed view faces counterevidence coming from a varied range of phenomena. Contemporary philosophers, logicians and linguistics have put the methodology to work in order to account for a great number of, apparently dissimilar, problematic cases. In this section I will consider a well-known class of cases and focus on the compositionally closed account that has been offered. This will help us understand better the explanatory performance of the closed view and whether it is well motivated. There are at least three different classes of problematic cases against compositional accounts. These are commonly known as cases of *context-of-use sensititvity*, *linguistic-context sensitivity* and *meaning flexibility*, particularly of proper names. I will consider the first of these classes, as it will help illustrate the limits and use of the closed view—for details on the other classes see Janssen (1997) and Szabó (2012) and (2013).

As Janssen (1997) points out, compositionally closed accounts of problematic phenomena follow either one of three general strategies. They either postulate *new meanings*—introducing a new parameter or a new function in the algebra—add *new expressions*—creating homonyms with different meanings that the syntax may select—or posit *new constructions*—imposing more syntactic structure and their accompanying syntactic rules. Consider the case of sensitivity to the context of use. The closed view has problems accounting for expressions whose meaning changes according to the context in which they are used (e.g., "I," "you," "here," "there"). When Richard uses "I," "I" contributes "Richard" as meaning. Such meaning is not part of the lexicon

and is not given to the English pronoun "I" in isolation. Furthermore, the meaning of the complex expressions in which it appears also varies depending on context. "I authored *Meaning and Necessity*" means something when Rudolf says it and something different when Richard says it. Thus, the meaning of the complex expression "I authored *Meaning and Necessity*" is not determined in accordance to *strong compositionality*.

The most accepted account of these phenomena is of course Kaplan's (1989). David Kaplan was perhaps the last disciple of Carnap, and had the rare opportunity of working with both Carnap and Montague. As such, Kaplan (1989) is a paradigmatic example of how to use logic to offer a full mathematical account of natural language, or a fragment of it. Kaplan (1989) is intended as a complete, compositionally closed, account of expressions whose meaning is sensitive to the context of use, including of course an account of the logic needed to make things fit properly in an algebra for all complex expressions. The resulting account is complex and impressive, but it is based on a rather simple strategy. Kaplan (1989) posits *new meanings*, including both parameters and functions from these parameters to the relevant values, to account for the missing meanings that appear to change from context to context.

According to this view, expressions that shift their referent from context to context are to be considered special, usually known as *indexicals*. These peculiar expressions do have a meaning in isolation. This meaning is not the one that gets to be part of the meaning of a complex expression, but it does determine, relative to a context, what gets to be that meaning. As such, expressions that are sensitive to the context of use get to have two *kinds* of meaning. On the one hand, they have a fixed meaning, one they have so to speak *in isolation*, and can be said to be part of the lexicon. This meaning is typically understood as a function ranging from parameters (e.g., a speaker, a location, a time, a world, etc.) to elements of the context of use (e.g., the speaker, the time of the utterance, the location of the utterance, etc.). On the other hand, these expressions have a changing meaning, one they have only *relative to a context*. This meaning is fixed by the function as the value of the parameter in a given context. Thus, "I" has as fixed meaning a function that picks out *the speaker* from the context of use, and it has as its context-relative meaning whatever individual that function picks out in that context. Thus, the function associated to "I" picks out an individual when Rudolf says "I authored *Meaning and Necessity*' and a different individual when Richard says it. This explains why "I authored *Meaning and Necessity*" has different meanings depending on who says it, and it does so in a fully compositional manner. The meaning of the complex expression "I authored . . ." is fully determined by the meaning of its parts, including the function associated to "I" and its context-relative value.

There are two reasons that speak on behalf of Kaplan's use of the *new meanings* strategy. First, the strategy succeeds in offering a compositionally closed account of problematic expressions. Second, and most importantly, the strategy is well supported by *independent evidence*. Empirical studies on the acquisition of indexical and demonstrative expressions support the view, as competence with such expressions is known to presuppose a host of capacities including intention understanding and contextual resolution from early on (see Salomo and Liszkowsky 2013; Nadig and Sedivy 2002; and Hanna, Tanenhaus, and Trueswell 2003). This makes Kaplan's positing of *new meanings* a well-supported auxiliary hypothesis (see Roberts 2002; and Elbourne 2008; for alternative accounts).

It should be obvious that this strategy can be extended to offer compositionally closed accounts of (almost) *any* problematic expression whose meaning appears to shift from context to context. In fact, Frege (1892) inaugurated the strategy (so to speak) in his effort to account for the otherwise problematic behavior of ordinary proper names. Frege (1892) notes that proper names that are coreferential, and so, intuitively synonymous may nevertheless be used to convey different contents. The statements "Mark Twain authored *Huckleberry Finn*" and "Samuel Clemens authored *Huckleberry Finn*" do not seem to convey the same meaning, even though the names "Mark Twain" and "Samuel Clemens" are coreferential. Accepting that synonymous proper names may be used to convey different contents is equivalent to rejecting strong compositionality as the statements in question would differ in complex meaning without differing in the meaning of the expression parts or in the way these are combined. To avoid this damning result, Frege (1892) famously postulates the existence of a second semantic level, beside the referent, for ordinary proper names. Further discussion concerning proper names has shown, if anything, that ordinary proper names are much more problematic than Frege (1892) suggests (see Kripke 1980; Soames 2002; and García-Ramírez and Shatz 2011). If we want to keep a compositionally closed account of ordinary proper names in natural languages we may not only need *new meanings* but also *new constructions*, turning proper names into syntactically complex expressions that may behave as both proper and common nouns (see Matushansky 2008).

Yet, unlike the case of "I," "here," and "now," the postulation of *new meanings* (and maybe also *new syntactic structure*) for ordinary proper names is *not* supported by evidence independent of any compositionally closed motivation. Ordinary proper names, such as "Mark Twain" are *not* obviously context-sensitive. If anything, they seem intuitively to be labels attached to individuals (see Kripke 1980). For example, independently of how it is used, "Mark Twain" appears to be just a label for Mark Twain. This makes the *new meanings* (and maybe also *new constructions*) strategy an auxiliary

hypothesis that is *not well supported*; even if it succeeds in offering a compositionally closed account of ordinary proper names, this strategy appears to be *ad hoc* because the main reason one has to posit *new meanings* in the case of proper names is, ultimately, the assumption that natural languages must observe *strong compositionality*.

The fact that *new meanings* strategies have been used to offer compositionally closed accounts of a rather wide range of seemingly unrelated expressions is evidence of how extensively accepted (and deeply assumed) is *strong compositionality*. Whenever there is some evidence that an expression appears to change its meaning relative to a context, the closed view will automatically resort to positing *new meanings* (and/or *new constructions*) to avoid counterexamples. Lewis (1996) argues for a similar account with respect to "knows" in ascriptions such as "John knows that it will rain tomorrow." Stanley and Szabó (2000) offer a related account for quantifier expressions "every," "some," "all," "no," "most," "many." In fact, the list of context-sensitive expressions appears to be open and growing, it now includes gradable adjectives such as "tall," and color terms such as "green" (see Preyer and Peter 2005; Stanley 2007; García-Carpintero and Kölbel 2008; and Silk 2016; for discussion see Bach 2005).

It is perhaps uncontroversial that these implementations of the *new meanings* strategy result in compositionally closed accounts. One can safely assume that the resulting accounts find a way to include the needed meaning for the problematic expression as just another part of the algebra for a given natural language. However, as it happens with ordinary proper names, it is far from clear that all these accounts are independently well supported. Another well-known case, to my mind parallel to that of proper names, is that of modal expressions. Kratzer (1977) and (1981) famously posits new meanings—i.e., a fixed function taking two arguments, a modal restriction and a scope—for expressions such as "must," "can," "necessary," "contingently," "possibly," and "might." Doing so allows Kratzer (2012) to offer a single fixed meaning for, say, "must" that accounts for its different meanings such as the deontic one in "Philosophers *must* study mathematical logic," the epistemic one in "If natural language is compositionally closed, modal expressions *must* have a common underlying meaning," and bouletic in "The mathematical account of natural language *must* be the best one." Yet, it is hard to find *independent* evidence for the auxiliary hypotheses that (following Kratzer 2012) modal expressions have a fixed associated function that takes as arguments a modal base and a proposition. If we are not looking for a mathematical account of modal expressions, it seems easier to simply accept that modal expressions do not behave in a compositional-friendly manner.

Without sufficient independent support, extending the *new meanings* (and/or new constructions) strategy across the board, based on the claim that it produces compositionally closed accounts, delivers unsupported extraneous semantic and/or syntactic hypotheses. This becomes clear once we recall the original dialectic. The closed view assumes that natural languages observe *strong compositionality*. Ordinary use of certain expressions, proper names and modal expressions among them, appears to contradict this assumption. Applying the *new meanings* and / or *new constructions* strategy to accommodate all problematic expressions is tantamount to positing auxiliary hypotheses designed to save the closed view from being falsified. Unless there are independent reasons to buy into the complex semantics and syntax of the closed view, all such accounts appear to be *ad hoc*.

Proponents of the closed view do not, of course, consider their accounts to be *ad hoc* or their hypotheses to be extraneous. This is not merely because they *are* looking for a mathematical / compositionally closed account but, most importantly, because they assume that any satisfactory account *should* be looking for a mathematical / compositionally closed explanation. This assumption is commonly supported by three different items of independent empirical evidence, and a theoretical observation. On the empirical side, it is argued that we need *strong compositionality* to account for the *productivity*, *systematicity*, and *simplicity* of natural languages. On the theoretical side, it is argued that assuming *strong compositionality* provides the best methodology when it comes to explaining how complex expressions get their meaning. Thus, even if there is no independent support for each compositional account of a given expression, there certainly is independent support for the assumption that ultimately motivates such accounts, or so it is argued. In the following section I will consider the aforementioned independent support for *strong compositionality*. As I hope to make clear, the demand for sufficient independent support will be left wanting.

2.4 INDEPENDENT EVIDENCE FOR THE CLOSED VIEW

Explicit arguments on behalf of *strong compositionality* are rare. Most theorists working within the closed view simply assume that the principle holds either as an empirical claim or as a methodological one. Still, there are at least four different arguments (see Szabó 2012 for more; see also Janssen 1997; Fodor 2001, and Baggio, van Lambalgen, and Hagoort 2012) that are widely accepted as offering the strongest support for the view. On the empirical side we have the arguments from *productivity*, *systematicity*, and *computability*,

and on the methodological side we have an argument based on the theoretical virtues associated to the principle. Let us consider all four of them.

Productivity

The number of distinct complex expressions that can be constructed within any natural language is potentially infinite. In this sense, there is no limit to the expressive *productivity* of natural languages. Still, humans with finite cognitive capacities manage to acquire and develop natural languages rather easily, eventually being capable of understanding every possible expression of the relevant language. There seems to be a tension between the *productivity* of natural languages and the fact that cognitively *limited* beings have all the necessary means to fully understand them.

Productivity poses a challenge to any theory of language, namely, to explain how limited humans can understand an unlimited amount of complex expressions. Humans cannot store an unlimited amount of complex expressions in memory in order to retrieve them when needed. It seems obvious, then, that knowing a language £ is *not* a matter of actually knowing all the possible expressions of £ and their meanings, but of being able to determine the meaning for any of such expressions. There must be some finitely describable knowledge that, once acquired, allows humans to understand an unlimited set of complex expressions and their meanings. *Strong compositionality*, it is argued, offers the best answer to this explanatory challenge. Recall the principle:

Strong Compositionality: The meaning of a compound expression *E* of £ is completely determined by the meanings the expression parts $<e_1, e_2, \ldots e_n>$ of *E* have in isolation in £, and the syntactic way the expressions are combined.

According to *strong compositionality* three elements are necessary and sufficient to determine the meaning of any complex expression *E* of a language £. We need, first, a lexicon including all the simple expressions *e* of £ and their meaning in isolation. We need, second, recursive syntactic rules of combination and, third, we need the corresponding recursive semantic rules of combination. The lexicon is finite, and so is the set of syntactic and semantic recursive rules of combination. Thus, if £ in fact observes *strong compositionality*, then there is a simple mechanical procedure to determine the meaning for every possible complex expression *E* of £, and all that humans need to do to master it is to acquire all three elements above described, all of which are finite and will typically make minor demands on memory and cognition.

Thus, the fact that natural languages are productive seems to offer independent support on behalf of *strong compositionality*. Exactly what kind of

support this is, and what kind of argument we can derive from it, is not obvious. It is clear that, by endorsing *strong compositionality*, the closed view does have a straightforward explanation of *productivity*. But it is unclear whether the latter is evidence of *strong compositionality* or something else. As a matter of fact, *productivity* is insufficient evidence for *strong compositionality*. To see this compare what *productivity* requires against what *strong compositionality* offers.

Productivity demands an explanation of how limited humans can acquire an unlimited understanding. The challenge is met once we describe a finite procedure by means of which limited beings may determine the meaning for an unlimited number of complex expressions of a given language. Nothing is said about what this procedure should be like. Whether it is purely algebraic, purely pragmatic, or if it involves a mix of syntactic, semantic, pragmatic, and perhaps other higher order cognitive human resources, is yet to be determined. However, *strong compositionality* has some such procedure on offer. It is purely algebraic in nature, and excludes the use of any non-syntactic or non-semantic processing. Thus, by endorsing *strong compositionality,* the closed view offers an account of *productivity*, but it also offers much more, such as an algebraic and logically closed procedure for determining the meaning of any possible complex expression. It is these *extra* requirements that are highly problematic; they lack independent empirical support and their observance gives place to unorthodox syntax (Janssen 1997) or semantics (Zadrozny 1994) that may, paradoxically, end up placing more cognitive demands upon an already limited form of cognition.

So *productivity* does not constitute independent evidence of *strong compositionality*. To illustrate this point consider the *weak* version of the principle of compositionality, one that is open to there being non-mathematical procedures that determine the meaning of complex expressions.

Weak Compositionality: The meaning of a compound expression is a function of the meanings of its parts and of the way they are syntactically combined. (Partee 1984, 281)

Suppose, further, that we *weaken* more this already weak principle by pushing it further away from any algebraic interpretation. The resulting principle would state something like

Weaker Compositionality: The meaning of a compound expression is *partly* a function of the meanings of its parts and of the way they are syntactically combined, and *partly* a function of *domain general* influences and other relevant human cognitive limitations.

There is little more one can do to offer a non-algebraic account of the meaning of complex expressions. It is not possible to offer an algebra that includes the relevant pragmatic and other cognitive requirements, because there is no way to predetermine which of these will be relevant. Determining which influences are relevant in a case will depend on which is the best (most charitable, rational, etc.) interpretation of the speaker. In fact, according to *weaker compositionality*, one and the same complex expression *E* may have its meaning determined by different pragmatic and other cognitive influences in different contexts. Clearly, *weaker compositionality* differs substantially from *strong compositionality*. It is not a principle that may be endorsed by the *closed view*. Yet, *weaker compositionality* has all the needed elements to account for *productivity*. It comes with a finite lexicon, and a finite set of syntactic and semantic recursive rules. It also requires some pragmatic and other cognitive abilities, all of which are possessed by humans. These elements do not deliver an algebraic procedure, but they do offer a finite procedure by means of which humans can determine the meaning of any complex expression they may encounter. Nothing more is needed.

What kind of argument can we derive from *productivity* on behalf of *strong compositionality*? There is no quick answer to this question. As far as I know, the fiercest case one can make is to claim that *strong compositionality* offers the *best* (though not the only) explanation of the evidence. Thus far, no one has tried to offer such a defense (see Szabó 2012). It seems, then, that *productivity* does not offer a clear and strong support for *strong compositionality*. Something similar happens with the remaining two empirical arguments for the principle, based on the *systematicity* and *computability* of natural languages.

Systematicity

Natural languages are known to be *systematic*. The meaning of some complex expressions appears to be related to the meaning of *other* complex expressions so that whoever understands the meaning of the former may also understand the meaning of the latter. Thus, for example, whoever understands the meaning of "Mary loves John" will also understand the meaning of "John loves Mary." Generally speaking (see Baggio, et.al. 2012), whoever understands expressions E_1 and E_2, made up with distinct expression parts but identical syntactic combination, will be able to understand any other expression $E_3 \ldots E_n$ made up from any of the expression parts of E_1 and E_2 and the same syntactic combination.

The explanatory challenge posed by the *systematicity* of language is different from that of *productivity*. There is no tension between the limits of language and those of human cognition. Instead there is evidence that simple

expressions (or expression parts) and syntactic combinations play a substantial role in determining the meaning of complex expressions. To account for *systematicity* is to explain how syntactic combinations and the meaning of simple expressions manage to determine the meaning of complex expressions in such systematic way.

Now, according to *strong compositionality* all simple expressions have a meaning in isolation and the recursive syntactic rules fully determine how these may be combined to form complex expressions and their meanings. If natural languages observe *strong compositionality*, then the meaning of *all* complex expressions is determined by the meaning of the simple expressions and the way they are syntactically combined. On this view, systematicity is not a surprise but an expected consequence of the strong compositionality of natural languages. To put it somehow, *systematicity* turns out to be a reflection of the algebraic compositionality that natural languages have according to the closed view. It seems, then, that *systematicity* supports *strong compositionality* more sufficiently, as it seems to demand something as specific as the principle offers. How else could we account for the *systematicity* of natural language if not by means of an *algebraic* account?

Once again, it is clear that *strong compositionality* directly explains *systematicity*. But it is not so clear that *systematicity* is evidence of *strong compositionality* and not, perhaps, something weaker. What *systematicity* requires is an account of meaning for complex expressions in which the meaning of simple expressions and their syntactic combination play a substantial and systematic role. Yet, a substantial and systematic role for lexical meanings and syntactic combinations is still far from being equivalent to an algebraic account. This is evidenced by the fact that *weaker compositionality* also offers a direct account of *systematicity*. According to *weaker compositionality*, we have the same building blocks, so to speak, as we do with *strong compositionality*, except that they are not all that matters. To fully determine the meaning of any possible complex expression we need more elements, including pragmatic and other cognitive influences. If natural languages observe *weaker compositionality*, then the meaning of several (perhaps most) complex expressions is partly determined by the meaning of simple expressions and their syntactic combination. This is a substantial and systematic role. Since syntactic combinations and lexical meanings are always determinant in the same way, *systematicity* becomes less mysterious. The fact that there are extra elements does not go against this substantial and systematic relation.

Briefly put, the kind of *systematicity* that can be observed in natural languages is too weak to support the kind of *algebraic* systematicity that *strong compositionality* predicts. As such, *systematicity* does not truly offer independent empirical evidence for *strong compositionality*.

Computability

More recently a new empirical argument has been offered on behalf of *strong compositionality*. Owed to Pagin (2012) this argument is based on the *learnability* of natural languages, i.e., that fact that natural languages can be acquired by limited humans. This argument differs from the argument from *productivity* in virtue of focusing on the ease of comprehension of natural languages by competent speakers and not so much on the tension between the unlimited *productivity* of language and limited human cognition. From this point of view, in order to explain how limited humans can learn natural languages we must offer an account of the *computability* of the latter.

Naturally, *strong compositionality* is itself an account of the relevant sort as it describes exactly how it is that natural language is computed under limited resources. The algebra that follows from observing *strong compositionality* is computable by human beings, or so it is alleged. Of course, as we have already seen with *productivity* and *systematicity*, *computability* is too weak to support *strong compositionality*. Pagin (2012) is aware of this and notes "a semantics may be computable without being compositional (and vice versa)." (Pagin 2012, 510) Thus, we need more than just *computability* if we want to defend *strong compositionality*. Pagin thinks we can find this extra element by looking into linguistic communication and, in particular, "at the feature that we manage to convey new contents by means of new sentences *in real time*, that is, that a hearer manages to compute the meaning online of an uttered sentence at a speed that matches the speed of speech." (Ibidem)

Pagin suspects that *real time computability* is possible only if the semantic processing involved requires relatively few and relatively easy steps. This seems, intuitively, correct. One can follow Pagin and claim that, when picking between alternative semantics, the one that offers the simplest computation is the better one. So far so good, but how does this get us into *strong compositionality*? If we want to seriously consider *real time computability* as evidence of something about natural languages we must, first, identify all the relevant cognitive resources that humans in fact have and put to use when *processing* speech. Just as it seems obvious that a simpler semantics is a more computable one, it seems obvious that the measure of processing simplicity required for natural languages is determined in relation to human cognition. In other words, the fact that humans process speech *in real time* is not evidence of natural languages having a simple semantics *simpliciter*; what is evidenced is the fact that natural languages have an interpretation procedure that is simple enough *for humans*—whether it is compositionally closed is yet to be determined.

If the *computability* argument is to succeed two important requirements must be met. First, we need to determine a measure of *psychological*

complexity that can help us compare among different accounts in terms of their *psychological simplicity*, which is the relevant sense of *simplicity* to consider. Pagin (2012) is partly aware of this. After considering three types of complexity measures—time, space, and Kolmogorov complexities—from computational theory, Pagin proposes that we measure the size of the representations of contents of the complex expressions and then define the complexity of the computation relative to the number of operations of Turing machines needed to process such representations. The result is, of course, a very clear notion of computational complexity. Whether this measure is substantially related to the relevant measure of human psychological complexity is still an open question.

Suppose, however, that we somehow find such relevant measure, the argument would still face another problem. As Szabó (2012) points out with respect to the *productivity* argument, the claim that natural languages are simple enough for *real time computability* is based on examples of complex expressions that hearers *already* understand. The fact that hearers understand complex expressions in real time "shows nothing more than that the information necessary to determine what they mean is available (to the hearer) immediately *after* they have been uttered. If there are features of the context of utterance that (the hearer) can invariably rely on, those features may well play a role in interpreting (those) complex expressions." (Szabó 2012, 76) Briefly put, the computability argument supports, at best, the weak claim that the interpretation of complex expressions of natural languages is simple enough for human beings to understand them in real time. Yet, the process of interpretation may involve linguistic as well as non-linguistic cognitive processing that, thanks to cognitive psychology and psycholinguistics, we know humans are capable of exercising in real time.[2] If so, then *strong compositionality* does not follow from the *real time computability* of natural languages.

It may well be, as Pagin (2012) argues, that compositionally closed languages are more computationally simple than languages that simply observe *weaker compositionality*. Nonetheless, it may also be that languages that observe *weaker compositionality* are simple *enough* for humans to engage in real time linguistic interpretation. The only way to adjudicate upon this issue is by means of empirical inquiry. We know for sure that natural languages are computationally simple enough for humans, now we need to go and find out which resources humans actually exercise as part of such computation. Fortunately for philosophers and linguists, cognitive psychologists and psycholinguists have been addressing this question for decades. I believe we now have enough empirical evidence to resolve the case against *strong compositionality*, or so I will argue in the following chapter. In any case, as

Pagin (2012) admits, "no strict argument for compositionality is forthcoming," at least not from the *real time* computability of natural languages. (Pagin 2012, 528)

Methodology

According to Janssen (1997), *strong compositionality* should be viewed as a methodology, perhaps the best one when it comes to accounting for the meaning of complex expressions in natural language. As such, *strong compositionality* tells us how to solve semantic problems, either by adding new meanings, new syntactic structure, or both. Furthermore, it helps "find weak spots in non-compositional proposals." (Janssen 1997, 461) This simple and clear methodology is, of course, generalizable. One can modify the semantics or the syntax (or both) as needed to account for as many puzzling phenomena as required, so long as the resulting account observes *strong compositionality*.

The simplicity of the methodology is certainly extraordinary, as is its explanatory power. If the suggestion is correct, there is pretty much no problematic semantic phenomenon that one may fail to account for by following the steps suggested by *strong compositionality*. In fact, the methodology is *so* surprisingly powerful that it becomes suspicious. It seems as if, on this view, there are no constraints on linguistic theorizing except for the observance of *strong compositionality*. This, of course, is not the intended way to understand the methodology, for it would otherwise be trivially true that all languages (natural or formal) observe *strong compositionality*. If one can simply change the semantics and the syntax accordingly, then it is no surprise that one can give an appropriately compositional semantics. So viewed, the methodology is rather poor. It turns out to be more like a practical methodology for language design than a theoretical methodology for understanding natural language. It must be, then, that this simplistic methodological reading is not the appropriate one.

A more appropriate methodological use of *strong compositionality* would require further, independent restrictions. To begin with, independently of the account one wishes to offer, the syntax and semantics of the relevant language should be fully specified. It is then an open question whether one can give a compositional account of the relevant phenomenon, given such syntax and semantics. The methodology suggests one can change either the semantics by adding new meanings, the syntax by adding new structure, or the lexicon by adding new expression parts. Yet, proceeding in such a way is a much more complex task than one would have thought. For now we have an independently specified syntax and semantics, and any changes will require independent motivation. Merely claiming that the proposed changes deliver a compositional account will not do.

It is, of course, not at all trivial to find a compositional account that observes all these independent restrictions. However, the methodology is no longer simply and straightforwardly suggested by *strong compositionality*. Before proceeding to offer a compositional account we must have an idea of the syntax and semantics of the relevant language, and there is no way to do so without engaging in an empirical inquiry into actual language acquisition and development. Thus, the question of whether a given phenomenon can be offered a compositional account is no longer a simple methodological one, but a complex empirical issue. Thus, it is not obvious that a compositionally closed methodology can in fact be extended to account for *any* problematic semantic phenomenon.

To sum up, *strong compositionality* only offers a simpler methodology if it is allowed an unrestricted leeway. But then the resulting account(s) will be trivially true. If the methodology is, as it should be, given proper independent restrictions, then it will turn out to be a more complex, piecemeal methodology in need of independent empirical evidence. Once we reach such piecemeal, empirically based methodology there is no clear advantage of *strong* over *weaker compositionality*, which relies on a case by case, empirical procedure. This will be made clear in chapter 4, where I will present the view I call *open compositionality* and its claim that natural languages are, at best, *weakly compositional*. Furthermore, as soon as we realize that the question about the compositional nature of human languages is an empirical one, it is no longer clear what the right answer is. If the resources and information that actually take place in human understanding of language are only a matter of syntax and semantics the closed account will work. We can always add (or take) more elements to (or from) the algorithm to get the proper, computable algebra. Something different happens if we find non-linguistic knowledge playing a substantial role in human understanding of complex expressions. If this is the case, the relevant information and processing will not be an extra element we can add to the algorithm. Rather, we would be facing instances of language understanding that are open to influence by non-compositional processing. In chapter 3, I will argue that this is in fact the case.

2.5 COGNITIVE PSYCHOLOGICAL FACTS AND SEMANTIC THEORIZING

I have argued that without any independent restrictions *strong compositionality* is bound to be trivial. If the theory is to be serious about its attempt to account for human languages, then the nature of human cognition will itself provide the much-needed restrictions for *strong compositionality* to constitute a meaningful methodology and/or empirical hypothesis. As of today,

our best approach to human cognition and language acquisition is offered by cognitive psychology and psycholinguistics (among other cognitive sciences). Not surprisingly, philosophers, logicians, and linguists endorsing the closed view have explicitly rejected the relevance of psychology for the study of how complex expressions get their meanings in natural languages. Some merely state such rejection; others have tried to offer some arguments. According to Thomason (1974), Montague considers "the syntax, semantics, and pragmatics of natural languages are branches of mathematics, not of psychology." (p.2)

Partee (1982) and (1988), for example, argues that a complete compositionally closed semantics for natural languages is impossible, given the flexibility of the latter. As I have shown, *strong compositionality* demands that simple expressions have a fixed meaning in isolation. Yet, Partee (1982) and (1988) notes, there are many expressions whose meaning is fixed either between speakers and the world (e.g., proper names and natural kind terms), between speakers themselves (e.g., compound expressions), and even some for which there is no correct interpretation (e.g., theory-dependent expressions such as "semantics"). Still, her doubts about *strong compositionality* aside, Partee (1988) appears to reject the relevance of psychology for natural language semantics as part of her criticism of Schiffer's (1987) argument that there can be no theory of meaning. However, the notion of psychology used by both, Partee (1988) and Schiffer (1987), has little to do with the cognitive scientific disciplines and more to do with what is known as folk psychology. Schiffer (1987) appeals to psychological facts in his attack against compositional semantics, those facts being beliefs, desires, and hopes. Partee (1988) argues that semantic facts, facts about how the meaning of certain expressions is determined, should be kept apart from these psychological facts. Clearly, Partee's (1988) claim is compatible with the acceptance of the relevance of other kinds of psychological facts, call them "cognitive facts," for our theoretical explanation of natural language processing and understanding. Like Partee (1988), cognitive psychologists and psycholinguists do not seem to consider what speakers believe, desire, and hope as relevant to determine which resources they use when acquiring, using or understanding the use of a natural language. Cognitive facts, facts about the resources deployed by human beings, should be distinguished from folk psychological facts, facts about the beliefs, desires, and hopes of human beings. Once we make this distinction, it is clear that Partee's (1988) argument against the relevance of folk psychological facts to understand natural languages cannot be extended to apply to cognitive psychological facts.

Janssen (1997) makes use of Partee's (1988) argument in order to drive a deeper gap between natural language semantics and psychology. According to Janssen (1997) Partee's argument shows that "a theory of natural

language semantics should be distinguished from a theory of natural language understanding." (Janssen 1997, 447) As with "psychological facts," "understanding" can be taken to mean two different things. It can have a folk-psychological interpretation, where natural language understanding involves whichever folk-psychological mental states of speakers are formed in virtue of their competent use of natural language. Yet, "understanding" may also have a more cognitive-psychological meaning, where natural language understanding involves whichever cognitive resources are needed for a competent use of natural language. As I said before, Partee's argument is only successful if applied in the first sense. It is the second sense of "language understanding" that is at stake here, and there are at least two good arguments showing that it *is* relevant for natural language semantics.

The first argument for the relevance of cognitive psychological facts for natural language semantics comes from recognizing the relevance of actual language use for our purposes. Actual language use is central to natural language semantics, as *explanandum* as well as *explanans*. According to Dummett (1973) and (1993), actual language use is *the explanatory goal* of a theory of natural language semantics:

> 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, 92)

A theory of actual language use is a theory of linguistic competence. As Schiffer (2003) and (2006) argues, in order to have an adequate account of linguistic competence we must be able to identify the "information processing that underlies, and thus accounts for, the person's ability to understand utterances in her language." (Schiffer 2006, 281) Now, actual language use is not only relevant as an object of study. According to many others (see Lewis 1992; Stalnaker 1999; Kaplan 1989; see also Larson and Segal 1995) actual language use is also a *source of explanation* for natural language semantics, since "it is our use of language that somehow determines meaning." (Lewis 1992, 106)

To see how important cognitive psychological facts are for natural language semantics one need only observe that, as a matter of empirical fact, actual language use is made possible by the cognitive psychological endowment of humans. The way in which humans in fact use language directly depends on which cognitive resources they have and put to use for such purpose. Briefly put, given the centrality of actual language use, cognitive

psychological facts are also central to natural language semantics. As Larson and Segal (1995) clearly put it, "(h)uman languages are, after all, the products of human minds." It seems appropriate (to put it mildly) to care for properly understanding the human mind when trying to understand its linguistic products (see García-Ramírez and Shatz 2011).

A second, partly methodological argument for the relevance of cognitive psychological facts for semantic theorizing is one we are already familiar with. We have seen that, without any empirical restrictions, the methodology derived from *strong compositionality* delivers trivial results. Any language can be turned into a compositionally closed one if we are given unrestricted leeway. If we are to do some serious semantic theorizing, we must avoid proceeding as if in a vacuum. Semantic theorizing "proceeds best within a framework of assumptions about the nature of semantic facts that allows for some reasonably explicit methodology." (Segal 2001, 548) Cognitive psychology provides some such framework. As I argued in the previous chapter, cognitive psychological theorizing about natural language fits well within the cognitive approach of Chomsky (1986) and Larson and Segal (1995), by assuming that semantic theorizing is in the business of accounting for human linguistic knowledge. Furthermore, by working within this framework we can also make sure that our account meets a theoretical compromise that is still little known even though it is presented by Lewis (1975). According to this compromise, we must guarantee that the language described by our theory is one that "could possibly—possibly in some appropriately strict sense—be used by a human population." (Lewis 1975, 171)

We are thus left with an open question. Can *closed compositionality* account for actual language use by describing a language that human minds can possibly—possibly in an empirically restricted and psychologically informed sense—acquire, develop, and use? In the following chapter I will argue that it cannot by presenting empirical evidence against *closed compositionality* or, in other words, against the claim that natural languages observe *strong compositionality*. If so then we must answer a further, perhaps more difficult, question: which alternative account should we offer? I will present an answer to this second question in chapter 4. This answer, which amounts to claiming that natural languages are compositional yet only partly and openly so, will be accompanied by an understanding of natural language semantics as a decision-making process—as opposed to a truth-condition-setting process—as well as by an alternative methodology of language based on the cognition-first hypothesis.

NOTES

1. On Montague's view (see Montague 1970a, and 1970c) English is not different from the artificial languages of logicians. His use of compositionality has been highly influential, with minor revisions on how to weaken the principle, how to make it compatible with contemporary advances in syntax and transformational grammar; and how to apply the proposal to account for discourse and not just isolated sentences. See Janssen (2012) and the essays in Partee (2004).

2. In fact, the computability argument faces a third important challenge. As Pagin (2012) strives to show, only "under certain restrictive conditions" do we get a semantics that is both minimally complex *and* compositional. Unfortunately, "these conditions tend not to be met in natural languages. There are reasons to suspect that syntactic complications and widespread context dependence make (the relevant kind of) compositionality impossible. Hence, in the end, no strict argument for compositionality is forthcoming." (Pagin 2012, 528)

Chapter 3

The Failure of Strong Compositionality

There is no meaning for complex expressions beyond what speakers take complex expressions to mean. As Larson and Segal (2001) put it, "[i]t is because English speakers take the strings of words *Camels have humps* to mean that camels have humps that those words have this meaning in English. If English speakers all took the sentence to mean that reptiles have wings, then this is what it would mean." (Larson and Segal 1995, 9) If we are to seriously consider *strong compositionality* as a hypothesis (or methodology) about the meaning of complex expressions in natural languages, we must find a way to determine if speakers in fact understand languages and their complex expressions in ways that are at least compatible with it. Thus, we need to test the empirical standing of *strong compositionality.* To properly do so we must first *operationalize* the principle. There are at least two ways in which the principle may be turned into an empirically testable hypothesis. First, it can be done by looking into the cognitive assumptions of *strong compositionality*, particularly with respect to cognitive architecture and its relation to language acquisition and development. These cognitive assumptions, in turn, may be understood as answering the question of what it is that humans must know, and which abilities they must possess, in order to learn and exercise a human language. Second, *strong compositionality* may also be turned into a processing claim, for instance, by looking into the principle's implications for language processing and comprehension (see Baggio, et.al. 2012). These implications may be seen as answering the question of how it is that competent speakers actually understand and process language.

In this chapter I will present and develop both ways to operationalize *strong compositionality*, as a set of cognitive assumptions in section 3.1, and as a processing principle in section 3.2. I will then consider the empirical evidence on language acquisition and development in sections 3.3 to 3.6; and

adult language processing in section 3.7. As I will show, the results are far from offering good news for the closed view, as they strongly suggest that *strong compositionality* is empirically false.

3.1 STRONG COMPOSITIONALITY AS A COGNITIVE HYPOTHESIS

The principle of *strong compositionality* claims that once we identify the lexical meanings, the syntactic combination, and the semantic rules associated with *any* complex expression we have thereby identified all the needed elements to determine *any* meaning it may have. No other information or knowledge is needed and no other may determine the expression's meaning. As Baggio et.al. (2012) argue, this *logical* closure entails that, prior to the complete determination of meaning for complex expressions, the information processing of natural languages is isolated from external influences, namely, from knowledge possessed by the speaker that does not constitute syntactic, semantic, or lexical knowledge. Put in terms of cognitive architecture (see Fodor 1983) natural language, like perception, exhibits *informational encapsulation*. In other words, the language system is "relatively impenetrable" to non-linguistic sources of knowledge (Baggio, et.al. 2012, 658). Jackendoff (1997) describes this in a rather simple and clear manner:

> The hypothesis of syntactically transparent semantic composition has the virtue of theoretical elegance and constraint. Its effect is to enable researchers to isolate the language capacity—including its contribution to semantics—from the rest of the mind, as befits the modular conception. It can therefore be seen as a potentially positive contribution to psychological theory. (Jackendoff 1997, 49)

To see the connection between the logically closed principle of *strong compositionality* and information encapsulation one need only notice that each of them entails the other. Assuming, with *strong compositionality*, that linguistic knowledge is fully (universal scope) constituted by lexical, syntactic, and semantic knowledge, then claiming that the knowledge needed to determine the meaning of any complex expression is *encapsulated* entails that no extra-linguistic (universal scope) knowledge may influence such determination. This is tantamount to saying that *only* lexical, syntactic, and semantic knowledge determines the meaning of any complex expression, which is what *strong compositionality* claims. Alternatively, claiming that natural language observes *strong compositionality* entails that the meaning of any complex expression is fully determined by linguistic knowledge

alone. This is equivalent to claiming that no extra-linguistic knowledge may play a role in determining the meaning of any complex expression, which is what information encapsulation claims with respect to language. In this sense, one can say that knowledge of language is *cognitively* closed, in the sense that it is not open to interact with other knowledge domains.

Now, cognitively speaking, information encapsulation requires a specific set of conditions for the encapsulated domain. In other words, for linguistic information to be encapsulated there must be a domain of knowledge or a set of cognitive abilities that are meant to work specifically for the linguistic domain. If this were not the case, if linguistic information were processed by domain general knowledge, then the linguistic domain would be penetrable by other domains by means of what is known as general knowledge. Thus, by endorsing *strong compositionality*, the closed view is committed to the claim that knowledge of language is a domain specific cognitive ability, which deals with encapsulated information. Knowledge of language, on this view, is not influenced by external sources and may have a one-way (inside-out) interaction with other sources only after the compositional process is concluded. Thus, *strong compositionality* is committed to the following principle of linguistic knowledge:

Domain Specificity: for any subject S, any complex expression *E* of any language £, S competently identifies the meaning *M* of *E* if and only if S fully determines *M* by means of language specific knowledge (i.e., lexical, syntactic, and semantic knowledge).

Aside from domain specificity, which determines how linguistic knowledge relates to other domains of cognition, *strong compositionality* is also committed to a further assumption about the internal structure of the domain, determining how the subdomains of linguistic knowledge (i.e., lexicon, syntax, and semantics) relate to each other. As we saw in the previous chapter (see section 2.2; see also Szabó 2012), for there to be *strong compositionality* the subdomains must be independent of each other. In other words, one subdomain should not influence or be influenced by another subdomain. This independence naturally derives from the *strengthening* needed to guarantee full compositionality (see 2.2). Lexical knowledge should be independent, since the meanings of compounding expressions must be given intrinsically or in isolation. These meanings are not determined by syntactic or semantic considerations pertaining to any complex expression in which they may appear. Syntactic knowledge must also be independent, for syntactic combinations must be fixed independently of the lexical meanings of any compounding expression they may be associated with; syntactic combinations settle the way in which those expressions may be combined. And if we want

to exclude the possibility of compounding expressions semantically combining in ways that do not correspond to the syntax of the compound expression, syntax must also be autonomous from considerations concerning semantic combination. Finally, semantic knowledge, in the sense of knowledge of rules for determining complex meaning, must be independent as well. These semantic rules are meant (ideally) to be functions going from syntactic combinations and lexical meanings into complex meanings. Syntactic considerations may help *select* which semantic rules apply in particular cases, but they cannot *determine* the rules that *constitute* semantic knowledge for a given language. If this were not the case, if syntactic or lexical considerations were to influence the semantic domain, then we would not be able to offer a simple algebra for all complex expressions; for each individual complex expression we would have to determine if and how lexical and syntactic considerations modify the semantic domain before we can go on selecting the relevant semantic combination rule and, only after, determining the meaning of the complex expression.

Hence, by endorsing *strong compositionality*, the closed view is also committed to the following principle of subdomain independence:

Independence: for any natural language £, £ will be constituted by a lexicon λ, a syntax σ, and a semantics ϕ if and only if λ, σ, and ϕ are independent subdomains of linguistic knowledge.

What must psychologically be the case for both, *domain specificity* and *independence* to be true? The answer is straightforward. First, if *domain specificity* is true, then there must be a set of language-specific abilities allowing humans to acquire, develop, and sustain knowledge of language, i.e., lexical, syntactic, and semantic knowledge. We can, thus, derive a more specific, empirically testable, hypothesis. I call this principle *cognition*.

Cognition: for any speaker *S* and any natural language £, *S* is a competent speaker of £ if and only if *S* possesses and/or exercises a set of language-specific abilities in virtue of which *S* acquires, develops and sustains her knowledge of £.

Second, if *independence* is true, then it must be that each subdomain of linguistic knowledge, whichever domain specific cognitive skill or knowledge it ends up being, is not itself the result of interaction with or among other subdomains. In other words, each subdomain must have autonomy with respect to the other subdomains. This gives us a more specific, empirically testable, hypothesis. I call this principle *autonomy*:

Autonomy: for any speaker *S*, any language £ with a lexicon λ, a syntax σ, and a semantics ϕ, *S* is a competent speaker of £ if and only if *S*'s knowledge of λ, σ, and ϕ do not interact amongst each other.

I claim that the evidence on language acquisition and development is incompatible with the cognitive assumptions of *strong compositionality*. The studies reviewed in section 3.3 show that competent speakers acquire, develop and sustain their knowledge of language by means of multiple, interacting, and domain general mechanisms or cognitive abilities, such as statistical analysis, intention understanding, and conceptual presuppositions. It follows, against *cognition* above, that there is no domain specific mechanism or cognitive ability responsible for acquiring, developing and sustaining competent speakers' knowledge of language and, consequently, no information encapsulation, as *strong compositionality* requires. Furthermore, the evidence also shows that the subdomains of linguistic knowledge substantially interact with each other, sometimes even bootstrapping each other from early on in acquisition and throughout development. It follows, against *autonomy*, that there is no subdomain independence amongst the lexicon, syntax, and semantics of natural languages. Thus, we can go back to our initial question, what is it that humans must know, and which abilities must they possess, in order to learn and exercise a human language? A quick answer, the details will come in 3.3, is that humans must possess a substantial knowledge of probabilities, statistical analysis, speaker intentions, and general conceptual knowledge. None of these are specifically linguistic forms of knowledge or information processing, yet they are capable of generating linguistic knowledge through interaction.

I said there is a second way to operationalize *strong compositionality* by looking into the principle's implications for language processing and comprehension. As I will show in the following section, these implications give place to testable predictions concerning how in fact adult speakers understand natural languages.

3.2 STRONG COMPOSITIONALITY AS A PROCESSING HYPOTHESIS

In the previous section I described what could be called the cognitive architecture of the linguistic domain that follows from accepting the logically closed principle of *strong compositionality*. Linguistic knowledge constitutes a specific domain, handling encapsulated information, and constituted by autonomous subdomains of knowledge with no substantial interaction amongst each

other. Let me now consider not the structure of linguistic knowledge but the exercise of this cognitive ability. How is it that competent speakers use their knowledge of language to understand complex expressions? Specifically with respect to *strong compositionality,* what kind of procedure do competent speakers follow when processing linguistic information? As I will show in what follows, *strong compositionality* does have detailed commitments concerning language processing. Not surprisingly, these deliver a logically *closed* procedure.

As we saw in chapter 2, section 2.2, the most common formulation of the principle of compositionality is in need of substantial strengthening if it is to meet its promise of offering a mechanical account of meaning for all possible complex expressions of a given language. The result is what I have been calling *strong compositionality*.

Strong Compositionality: the meaning of a compound expression *E* of £ is completely determined by the meanings the expression parts $<e_1, e_2, \ldots e_n>$ of *E* have in isolation in £, and the syntactic way the expressions are combined.

There are two substantial processing implications of *strong compositionality*. First, as I mentioned already, the function that completely determines the meaning of the relevant complex expression is a mathematical recursive one. Lewis (1975) makes this clear when describing a language (including natural languages) from the standpoint of the closed view. A language is, on this view,

> [A] function, a set of ordered pairs of strings and meanings. The entities in the domain of the function are certain finite sequences of types of vocal sounds, or of types of inscribable marks; if σ is in the domain of a language £, let us call σ a *sentence* of £. The entities in the range of the function are meanings: if σ is a sentence of £, let us call £(σ) the meaning of σ in £. What could a meaning of a sentence be? [. . .] a function from worlds to truth-values—or more simply, a set of worlds. We can say that a sentence is *true in* a language £ *at* a world *w* if and only if *w* belongs to the set of worlds £ (σ). (Lewis 1975, 163)

On the closed view an algorithm determines the meaning of any complex expression. In other words, the value of the meaning function—£(σ) in Lewis' terminology—is determined by a finite sequence of calculations fixed by *strong compositionality*. Given a complex expression as input, the hearer must (i) identify and process the expression parts and their lexical meanings, (ii) identify the syntactic combination of the former and, finally, (iii) apply the corresponding rules of combination, given the results in (i) and (ii), and deliver an output. Once the calculation reaches stage (iii) the meaning

determination process comes to an end, no further steps are needed or even accepted. This way of putting things makes it clear that, by endorsing *strong compositionality*, the closed view aims at offering an effective and mechanical way of determining the meaning of all possible complex expressions of a given language.

Thus, by endorsing *strong compositionality* the closed view is committed to the truth of *algorithm*:

Algorithm: for any competent speaker *S* and any complex expression *E* of any natural language £, *S* understands the meaning of *E* in £ if and only if *S* processes *E* following the algorithm given by *strong compositionality*—i.e., steps (i)-(iii).

This principle is committed to a very specific view of how competent speakers do in fact process speech. According to *algorithm* competent speakers follow a finite, self-contained, mechanical procedure when processing complex expressions of the relevant language. This algorithm describes a detailed path that has been portrayed as a bottom-up process (see Hintikka 1983), from the meaning of expression parts to the meaning of complex expressions by means of syntactic combination. No other direction is offered, and no backwards (top-down) direction is acceptable, since any such procedure would be equivalent to rejecting the closure of the compositional account. If, for example, once the interpreter reaches (iii) she goes back to revise the meanings offered in (i), then the independence of the lexicon would be infringed; the same happens if the interpreter goes back from (ii) to revise the results in (i), or from (iii) to revise the syntactic combination in (ii), this time disrupting the independence of the syntax.

Now, it is important to underscore an essential element of the compositionally closed proposal, having to do with the logically *closed* nature of the algorithmic processing proposal, namely, the completeness of its output. The dismissal of any backward step of re-calculation entails that *all* there is to the assignment of meaning to the complex expression is whatever the compositional procedure delivers. In other words, the unidirectional bottom-up procedure *always* delivers as output a complete and final assignment of meaning to the relevant complex expression. If this were not the case, if the relevant output could miss some crucial information, then the hearer would be justified to go back and revise the procedure *prior* to its conclusion. The phrase "prior to its conclusion" is of central importance. The principle of *strong compositionality* does not imply that no pragmatic revisions (or reinterpretations) are allowed. It merely entails that such revisions, if any there should be, must be *posterior* to the algorithmic procedure

and should be motivated only by pragmatic interests, not by any lack of meaningfulness in the algorithmic output. In any case, pragmatic reinterpretation cannot alter the algorithmic procedure by including non-linguistic information.

These considerations deliver a more detailed and empirically testable principle that follows from accepting *strong compositionality*, which I dub *processing*.

Processing: for any competent speaker *S* and any complex expression *E* of any natural language £, *S* understands the meaning of *E* in £ if and only if *S* processes *E* by following a bottom-up direction of algorithmic calculation, without any backwards (top-down) steps of re-calculation prior to the conclusion of the algorithmic procedure.

The processing *closure* of this principle should be evident. It purports to describe how competent speakers proceed in order to understand the meaning of any complex expression. The procedure is presented as a bottom-up and self-sufficient algorithm, it is *not open* to any external, domain general information or revision.

It is not difficult to see the relation between the cognitive commitments of *strong compositionality* and the *processing* principle. *Strong compositionality* is committed to both the encapsulation of linguistic knowledge and the independence of its subdomains. These commitments give place to the principles of *cognition* and *autonomy* specifying the cognitive architecture of linguistic knowledge. Now, if the exercise of such domain specific knowledge is to help determine the meaning of any complex expression, say while processing speech, it must be done in a way that observes both *autonomy* and *cognition*. That is precisely what the principle of *processing* guarantees. The output completeness of the algorithmic procedure guarantees that no external, or domain general, knowledge plays a meaning determination role. Furthermore, the dismissal of backward, top-bottom recalculation steps *prior* to conclusion guarantees the autonomy of each subdomain of linguistic knowledge.

As with *cognition* and *autonomy*, it is an empirical question whether competent speakers do in fact follow *processing*. Fortunately, there are multiple empirical studies on actual speech processing in adults that directly address this issue. As I will show in section 3.6, the resulting evidence does not support *processing*. In what follows, sections 3.3 to 3.5, I will consider various empirical studies on language acquisition and development, word learning, and syntax acquisition that are relevant to determining the empirical standing of *cognition* and *autonomy*.

3.3 LANGUAGE ACQUISITION AND DEVELOPMENT

As I have said repeatedly, by endorsing *strong compositionality* the closed view is committed to the truth of both *cognition* and *autonomy* (see section 3.1; see also below). In this section I will review empirical studies that will directly bear on these commitments. My goal is to present an overview of the relevant skills that are employed by subjects that acquire and use a natural language. Most studies are focused on the acquisition and development of lexical and syntactic knowledge. Semantic knowledge, understood as distinct from knowledge of the lexicon, becomes knowledge of the recursive semantic rules of composition for complex expressions. As such, it is considered to be a consequence of lexical and syntactic development. After carefully considering the empirical evidence on the nature of lexical and syntactic knowledge it will be sufficiently clear that the cognitive commitments of *strong compositionality* are lacking in empirical support. In the following section I will consider studies on the acquisition of the lexicon, and will look at studies on syntax afterwards. In this section I want to offer a clear view of the theoretical landscape concerning language acquisition and development in general.

The field of language acquisition and development is concerned with determining the nature of the cognitive means by which a human infant becomes a competent language user. To do so, students of language acquisition aim at determining, first, which are the prerequisites that a subject must satisfy—i.e, the knowledge and cognitive abilities an infant must have—in order to be capable of acquiring a human language. This set of cognitive prerequisites is commonly known as the "language acquisition device" (see Chomsky 1959; Shatz 2007a). A chief question to be answered is whether, prior to the acquisition of language, this language acquisition device is constituted by domain specific knowledge (e.g., an innate or prelinguistic knowledge of grammar) or if language can be acquired by a set of distinct domain general cognitive resources (e.g., intentional understanding, statistical learning, conceptual biases, etc.). Once the nature of the language acquisition device is determined, students of language must consider the question of change. How is it that such particular set of cognitive abilities (whether specific or general) develops in order to give place to adult-like use of language? Is there substantial, qualitative change or is it just a matter of maturation?

Theoretical approaches to answer these questions divide in two opposing groups (see Ambridge and Lieven 2011). To better understand this opposition it is important to have a clear view of the underlying dialectic. Any account of language acquisition is an account of the means by which humans get to an end state, i.e., adult grammar, from an initial state, i.e.,

infant knowledge. Most of the debate between alternative accounts concerns the proper characterization of the initial state. The more heavily endowed with special knowledge infants are, the easier it is to reach the end state. But there is also a debate concerning the end state itself. The more specialized the end state turns out to be the more knowledge it will demand from human infants at the initial state. And, of course, there is debate concerning the means, i.e., linguistic input, to acquire language. A richer linguistic input will reduce the demand of cognitive endowment at the initial state. The two opposing approaches to language acquisition are characterized by their claims on these dimensions. One view considers the end state a heavily specialized, formal, generative system that is unique to language cognition; and the input merely reflects the application of this formal system, and so is insufficient. Naturally, the initial state is required to be a heavily endowed one. The other view considers the end state as a set of non-generative rules, the applications of which are constrained by usage-based limitations. The input sufficiently exhibits these features and, hence, the initial state does not demand any heavy, domain specific, endowment. Let us now consider the details of each view.

On the one hand, universal / generative grammar approaches follow Chomsky (1959) in claiming that adult grammar is a highly abstract, specialized, and formal knowledge and, hence, take human infants to be endowed with a language acquisition device that consists of domain specific knowledge of such grammar. This grammar is universal, as it applies to and underlies the acquisition of all possible human languages, and it is generative, because it consists of highly general grammatical rules that apply to abstract linguistic categories and phrases—e.g., VERB, NOUN, VERB PHRASE, and NOUN PHRASE (see Ambridge and Lieven 2011, 1–6)—thereby fixing how grammatical structures are generated for any human language.

Generative approaches typically claim that such abstract and universal grammatical knowledge is needed for language acquisition given the impoverished nature of the input that an infant is exposed to (see section 1.2). Human infants, on this view, face the enormous task of acquiring a human language by being already equipped with substantial knowledge of grammar for that (or any other) language. This knowledge does change as subjects develop from infants to adults, but this change is better understood as a form of maturation, as knowledge of grammar gets refined into less general rules, this time applying to a single human language. Generative views are commonly associated with the further claim that knowledge of grammar is innate. Yet, as Ambridge and Lieven (2011) point out, it is possible to have a generativist approach that rejects the innateness claim. For current purposes it will be enough to consider such knowledge as prior to the onset of language acquisition (from now on "prelinguistic knowledge") while focusing on the

generativist's claims about the domain specificity of the relevant knowledge, its internal structure, and the autonomy of its parts as well as the nature of the change it undergoes through development.

On the other hand, socio-pragmatic accounts (Bruner 1975; Halliday 1975; Tomasello 1992 and 2003; Shatz 1981 and 1994; Diesendruck, Hall, and Graham 2006) reject the claim that adult grammar is a highly abstract system and, hence, deny that human infants are equipped with prelinguistic grammatical knowledge. On this view, adult grammar does not constitute a generative, formal system. Instead of being an abstract rule-based system that determines how grammatical structures are to be formed, adult grammar is considered to be a usage-based system, partially constrained by pragmatic restrictions. As a result, the initial state of language acquisition need not be heavily endowed. This is so not only because the end state is closer, but also because the input is taken to provide an enriched means to attain it. Linguistic input, it is claimed, is not as impoverished as generative accounts suggest (see Shatz 1987; Elman 2003; Hoff 2006; Salomo and Liszkowski 2013).

However, it is a mistake to think that socio-pragmatic accounts, in opposition to generative accounts, do not rely on prelinguistic knowledge of the language learner. As a matter of fact, socio-pragmatic accounts rely heavily on the idea that language learning requires the child to take an active role by bringing "a variety of capacities to the complex task of language acquisition, and based on whatever her capacities and knowledge at the time, she makes inferences about her linguistic situation at that moment." (Shatz 2007a, 9). Thus, socio-pragmatic theories also offer an account of something like a language acquisition device, which is conceived as the result of the exercise of domain general knowledge of different sorts. Language acquisition—i.e., the acquisition of grammar as well as of word learning—is the product of domain general knowledge, such as statistical learning, pattern recognition, and structure mapping (see Saffran 2001a, Saffran, Senghas, and Trueswell 2001; Lany and Saffran 2010), interacting amongst each other in coordination with an early understanding of mental states (see Bloom 2000; Baldwin and Moses 2001), such as referential intentionality (see Saffran, Aslin, and Newport 1996; Gentner, Holyoak, and Kokinov 2001), and some conceptual biases (see Markman 1992; Carey 2001; Spelke 1994). Thus, for socio-pragmatic accounts, there is no abstract knowledge of grammar. Grammar is *constructed* from the input through the interaction between multiple domain general cognitive systems. What is acquired through development is not a set of recursive rules that apply to abstract structures, but a set of structures that may be useful for communication. According to socio-pragmatic accounts, a basic desire to communicate and, thus, a basic understanding of intentions, attention, and use of language, plays a central role in the process of language acquisition.

Now, endorsing one or the other of these views on language acquisition and development will have different consequences with respect to the closed view and its cognitive commitments *cognition* and *autonomy*.

Cognition: for any speaker *S* and any natural language £, *S* is a competent speaker of £ if and only if *S* possesses and/or exercises a set of language-specific abilities in virtue of which *S* acquires, develops and sustains her knowledge of £.

Autonomy: for any speaker *S*, any language £ with a lexicon λ, a syntax σ, and a semantics ϕ, *S* is a competent speaker of £ if and only if *S*'s knowledge of λ, σ, and ϕ do not interact amongst each other.

Socio-pragmatic accounts of language acquisition directly contradict *cognition*, as they are based on the central idea that linguistic competence is *not* the product of any language specific cognitive system or domain of knowledge. Knowledge of language, on this view, is sustained by the interaction of multiple domain-general, socio-pragmatic, cognitive abilities. For socio-pragmatic theories, there simply is no such thing as a language-specific domain of knowledge, much less a closed system delivering meaning for every possible complex structure within the system. The same goes for *autonomy*, as the evidence on lexical and syntactic development shows, the multiple domains of knowledge involved in language acquisition appear to be in constant interaction amongst each other, as a larger lexicon helps acquire syntactic knowledge and syntactic knowledge helps acquire new words. Thus, briefly put, if socio-pragmatic accounts of language acquisition and development are true, the closed view of natural language is false.

What about universal/generative grammar approaches? Are they compatible with *cognition* and *autonomy*? Here the answer is less obvious. At first glance it seems as if the view is not only compatible but even motivates the principles. *Cognition* demands the existence of a language specific domain of knowledge. Generative grammar approaches to language acquisition claim that human infants possess a language specific knowledge of grammar, also known as the language acquisition device (LAD). The LAD directly satisfies *cognition* by offering empirical support. Or so it seems, that is, if in fact the LAD constitutes or is enough to give place to a compositionally closed system of representation. The LAD is explicitly described as knowledge of highly abstract grammatical rules. Thus, in and of itself, the LAD is far from constituting any system of representation. In fact, it is important that the LAD does not constitute any such system, since the latter are particular languages and the LAD is universal with respect to them. *Cognition*

demands a specific domain of knowledge that is already *too* specific for generative views, one that includes a lexicon, syntax and semantics of a particular language.

So the only way in which generative views may be compatible with *cognition* is if, as a result of the process of acquisition and development, the LAD maturates into a language specific, cognitively encapsulated, system with the shape and structure of a natural language as described by the closed view. Whether this is the case, of course, depends on the evidence and on the details of the generative grammar account of acquisition and development. There is, however, a rather clear pathway that the account should follow. If it is to be compatible with *cognition* and *autonomy* the maturation process should not allow for there to be influence of domain general cognitive abilities, or interaction among the subdomains of linguistic knowledge. Unfortunately, as I will show, neither the evidence nor the generative account satisfy these restrictions for at least two reasons. First, the evidence of extra-grammatical domains of knowledge influencing language acquisition and development is undeniable, even for generative accounts. Second, generative accounts aim at showing how adult *grammar* may result from maturation of the LAD, while semantic and lexical knowledge may require interaction with other more general domains. Briefly put, generative grammar views of language acquisition and development seem to be committed to the idea that knowledge *of grammar* is domain specific and perhaps encapsulated. But they do not seem to be so committed to claims concerning *natural languages*, or to claims of autonomy among the subdomains.

I believe both groups of accounts of language acquisition go against *cognition* and *autonomy*. If so, then these theories and the evidence they present constitute a strong empirical rebuke of the closed view. Let me, then, consider the evidence and accounts offered for word and syntax acquisition. To offer a somewhat detailed account of the evidence and debate I will be following the critical reviews of Saffran and Thiessen (2007), Baldwin and Meyer (2007), and Ambridge and Lieven (2011). The reader is advised to pay a closer look at these reviews for a thorough account.

3.4 WORD LEARNING

Natural languages such as English are among the most complex cognitive systems. Nevertheless, all human infants manage to acquire some or other natural language in a short period of time and with apparent ease. There is no doubt that there are innate cognitive abilities involved in achieving this feat. The debate concerns the nature of these abilities, whether they part of a

language-specific domain or if they constitute a set of domain general cognitive abilities.

The acquisition of word meanings constitutes a fruitful ground for this debate. At 18 months of age, human infants commonly have a 50-word lexicon; by their sixth year of age children possess an impressive lexicon of up to 10,000 words. How do they manage to do this? According to generative accounts, this impressive achievement is made possible thanks to language-specific innate abilities of word recognition, speech segmentation, and perception. There are two well-documented facts about word acquisition that have been adduced in support of this account, namely, fast-mapping and categorical perception.

There is fast-mapping whenever a subject can understand and competently produce a novel word after a single exposure to it. In a study with two-, three- and four-year-olds Heibeck and Markman (1987) found that human infants are capable of drawing a correspondence relation between a novel word and its referent even on the basis of a single exposure. Subjects in the experiment were presented with novel words (e.g., a color, shape, or texture term) by contrasting them with well-known words of the same domain. After a single exposure with the novel word, subjects were tested on their understanding and production of the novel word and of the semantic domain that corresponds to the word. The results show that even 2-year-old infants are capable of forming quick hypotheses about the novel word in order to successfully narrow down its meaning. This cognitive ability is known as fast-mapping. The study also shows that this fast-mapping of words with their meaning can also take place without linguistic contrast, so long as the context is compelling enough. In a more recent study with 4-year-olds, Waxman and Booth (2000) found differences between the use of these fast-mapping for learning words and for learning facts, suggesting that there may be at least principles involved in language learning that are not involved in other cognitive tasks.

Now consider categorical perception. The acoustic features of speech are, as a matter of physical fact, continuous. There are no gaps in between distinct sound patterns. There is categorical perception when this continuous stimulus is perceived as if it exhibited gaps, dividing the stimulus in what seem to be distinct categories. In the case of speech perception, it is interesting to notice that speakers exhibit categorical perception for phonemic contrasts. Liberman, Harris, Hoffman, and Griffiths (1957) show that adults are better at distinguishing between /b/ and /d/ contrasts than between different cases of /b/, even though the phonetic contrast between /b/ cases was phonetically as stark as that between /b/ and /d/. This result, however, does not seem to come up for non-speech phonetic perception (see Mattingly, Liberman, Syrdal, and Halwes 1971). Thus, the evidence strongly suggests that a special cognitive

mechanism is in place for language perception, one that may be particularly helpful when it comes to word learning.

In a study with 4-month-old infants, Eimas, Sigueland, Jusczyck, and Vigorito (1970) discovered that the subjects were able to discriminate between speech sounds that, for adult speakers, correspond to a distinction between voiced and voiceless stop consonants. For the /pa/ /ba/ contrast, adults identify /ba/ when onset of sound pattern takes less than 25 milliseconds, and /pa/ when onset takes more than that. The infants in the experiment exhibit adult-like performance. They are sensitive to phonetic contrasts that cut across the mentioned categories, but no such evidence was found for phonetically similar contrasts that did not do so. Further studies found evidence of categorical speech perception in 6-month-old infants for phonemic contrasts that correspond to categorical contrasts in languages other than their own (see Aslin, Pisoni, Hennessey, and Perey 1981). Together, these results strongly suggest that a speech-specific perceptual mechanism, if there is such, is innate.

Fast-mapping and categorical perception of speech seem to offer strong evidence on behalf of generative accounts of word learning. More important for our purposes, the evidence suggests that the cognitive commitments of the closed view are empirically supported, particularly *cognition*, as there seems to be evidence of a domain specific knowledge of language. However, further research has shown that the above-mentioned facts of word learning are far from constituting evidence of a special cognitive mechanism dedicated to word learning and language acquisition.

Fast-mapping has been shown to take place for non-linguistic tasks, such as fact learning. Markson and Bloom (1997) tested 4-year-olds and adults on their capacity to retain novel factual information about objects after a single exposure. The evidence shows that there is fast-mapping in domains other than the linguistic one, showing that fast-mapping is the result of cognitive capacities that are not circumscribed to the linguistic domain. Waxman and Booth (2000) argue that there is still a difference in the principles invoked in language learning that are not invoked in fact learning. But even accepting this is far from showing that there is an especially dedicated, language learning, cognitive mechanism. If, as Markson and Bloom (1997) suggest, fast-mapping is a domain general cognitive ability, it is not a surprise that its application follows different principles to solve tasks in different domains (see Saffran 2001a). Later research has shown evidence of fast-mapping in non-human mammals (see Kaminski, Call, and Fischer 2004), further supporting Markson and Bloom's domain general claim.

There is also mounting evidence that categorical perception is a domain general cognitive ability. It has been shown to exist in non-linguistic animals, and to be used by humans for non-linguistic purposes. Kuhl and Miller (1975) found categorical perception in non-mammalian animals, such as the

chinchilla; Kuhl and Padden (1982) consider evidence of categorical perception in macaques; Cutting and Rosner (1974) and Aslin and Pisoni (1980) present evidence of categorical perception for sounds that do not constitute speech (e.g., tones, music, etc.); Etcoff and Magee (1992) show that there is categorical perception for facial expressions; Bornstein, Kessen, and Weiskopf, (1976) found categorical perception in color perception; and Bornstein (1987) considers evidence of categorical perception in vision and audition in general.

It seems that neither fast-mapping nor categorical perception are language-specific cognitive abilities but, more likely, domain general cognitive abilities, perhaps constituting some of the fundamental features of cognition. These, of course, are bad news for generative grammar accounts, as well as for the closed view. The empirical evidence suggests there is no language-specific cognitive mechanism. Thus, there is no evidence of a highly abstract, innately inherited, language acquisition device or, alternatively for the closed view, of a domain specific knowledge of language. In fact, the evidential and theoretical ground appears to be much worse for these views, with some arguing not only that they are lacking in empirical support, but also that they cannot in principle offer an account of the evidence.

Theories of language acquisition and development that appeal to language-specific, innate, cognitive mechanisms are top-bottom accounts. That is, according to these views, language learners begin the task of language acquisition equipped with a set of language-specific categories—i.e, acoustic or syntactic. The task then consists of building an inventory—e.g., a phonetic inventory—by means of these pre-determined categories. Pierrehumbert (2003) argues that this task is impossible to achieve. She presents evidence from studies in phonetic typology showing an overwhelming language-specific acoustic variation in the production of phonemes that are taken to have the same linguistic features. If we try to build top-down a phonemic inventory, starting from a pre-determined set of acoustic categories, our inventory would mistakenly count distinct phonetic productions of the same phonetic distinction as phonetic realizations of distinct phonetic categories. Pierrehumbert (2003) convincingly argues that the only way to build a competent inventory is to do so bottom-up, from the overwhelmingly variable input to the phonemic categories. Pierrehumbert (2003) shows how this can be done by means of statistical analysis.

Like Pierrehumbert's (2003) account, socio-pragmatic theories of word learning rely on the availability and wealth of linguistic input. Infant directed talk (IDT) is rich in language learning information that substantially benefits the process of acquisition. IDT exaggerates prosodic characteristics of language, thus making phonological analysis easier for the learner (see Jusczyk 1997; Liu, Kuhl, Tsao 2003). IDT is also rich in emotional and intentional

content (Bruner 1983); and children are sensitive to this wealth (Fernald 1993). This extraordinary help from the input appears to be accompanied by a highly active learner, constantly seeking and sustaining social interaction with competent speakers. Based on detailed empirical research Shatz (1994) famously shows that infants' social understanding emerges early in life and develops gradually. This social competence has been shown to aid language acquisition and development. Young infants are highly sensitive to social clues that may help them attend to speaker's referential intentions (see Baldwin 1991; Hollich, Hirsh-Pasek, and Golinkoff 2000; Baldwin and Moses 2001; and Akhtar 2005), and so can be influenced by pragmatic and perceptual context (Diesendruck, Gelman, and Lebowitz 1998).

Not all socio-pragmatic accounts of word learning are centered on statistical analysis as heavily as Pierrehumbert's (2003). To acquire a novel word is chiefly a matter of identifying the meaning (or referent) that corresponds to the word. Yet finding the corresponding meaning may be a daunting task, since learners must pick among a huge variety of alternative hypotheses (see Quine 1960). Infants follow many different strategies that help them reduce such hypotheses into just a few from which to pick the corresponding meaning. Some theorists (see Baldwin 1991; Baldwin and Moses 2001) focus on how infants and young children interact socially and engage in joint attention, which gives them the cues that help them identify referential intentions and, hence, identify the corresponding object or set of objects. Infants are not only good at but seem to also be largely interested in determining what other people are thinking of. This ability to "read" other peoples' minds has been shown to play a substantial role in word learning (see Akthar and Tomasello 2000; Bloom 1998). Others have pointed out that children also benefit from syntactic knowledge of word classes. If a novel word is considered to be a proper name it is given a different treatment than if it is considered an adjective (see Hall, Waxman, Bredart, and Nicolay 2003). Markman (1989) famously suggests that infants follow three principled, yet revisable, biases in order to simplify word learning. Infants seem to assume that words typically refer to whole objects (as opposed to parts), that words commonly refer to kinds and not individuals, and that there is a one-one correspondence between words and objects or kinds. These assumptions are known as the whole-object, taxonomic, and mutual exclusivity biases. Other studies (see Carey 2001; Spelke 1994) suggest that human infants are innately equipped with knowledge of certain very general kinds. Prelinguistic infants seem to have an understanding of the distinction between individuated and non-individuated objects, between animate and inanimate objects, among others. This prelinguistic knowledge constitutes a conceptual bias that may guide the language learners in the process of word acquisition (see Soja, Carey, and Spelke 1991; Prasada, Ferenz, and Haskell 2002).

Taken as a whole, the evidence shows that language learners make use of *all* these strategies and cognitive abilities to successfully acquire, develop, and use a natural language. Knowledge from different domains—syntax, statistics, Theory of Mind, joint attention, lexical and conceptual biases—is needed to achieve such an extraordinary cognitive feat. Not surprisingly, integrative approaches to word learning—according to which multiple mechanisms are involved in language learning—are considered to offer the best explanation (see Hollich, Hirsh-Pasek, and Golinkoff 2000; Bloom 2000). Word learning is the product of the interaction between multiple cognitive abilities and knowledge from various (specific and general) domains. This is what the empirical evidence overwhelmingly shows, thus speaking against the cognitive assumptions of the closed view. There is no language-specific set of cognitive abilities, no specially dedicated mechanism, responsible for the acquisition or development of natural languages. Linguistic competence involves multiple abilities and domains of knowledge converge.

3.5 SYNTAX ACQUISITION

Skinner (1957) famously argued that language acquisition could be explained solely in terms of social interaction and reinforcement between infant and adult. Chomsky (1959) convincingly shows that this proposal is doomed to fail. As we saw in the previous chapters, natural languages are known to be productive. Competent speakers do not only understand previously heard sentences, but can also understand an unlimited number of complex expressions they have never heard before. A linguistic child is capable of understanding an unlimited set of new complex expressions, and this ability cannot be explained solely by previous social interaction and reinforcement, some structural (syntactic) knowledge of language is required.

The acquisition of syntax constitutes another fruitful ground for debate between socio-pragmatic and generative accounts. It is useful to recall (see section 3.3) the end-state / initial-state dialectic for this debate. Following Chomsky (1959), generative accounts take adult grammar (end-state) to be a highly abstract, specialized, and formal knowledge of sentence formation rules. Based on this idea, generative accounts typically make two further substantial claims. First, it is claimed that adult-like formal knowledge of grammar is too abstract to be learned from nothing. Second, it is also claimed that the linguistic input available in social interactions is too impoverished to constitute a learning ground for such formal knowledge. If so then language learners must come to the task of language acquisition already equipped with some knowledge of how natural languages are structured. Hence, infant

knowledge (initial-state) must already include some general knowledge of grammar.

There is little debate as to whether adult speakers possess an abstract grammatical knowledge, including structural knowledge that allows for sentence formation as well as knowledge of grammatical categories for different expressions. The generativists' first claim above is accepted by both parts of the debate; even socio-pragmatic accounts agree that language learners must start with something. What is disputed is the claim that this knowledge is domain-specific syntactic knowledge, such as a universal grammar, i.e., a set of highly abstract grammatical rules and principles that apply to all human languages. The heart of the controversy lies in the argument for a universal grammar, which stems from the generativists' second claim concerning the poverty of stimulus. Briefly put, generativist accounts of syntax acquisition claim that linguistic input radically underdetermines the syntactic principles and rules governing it, making it consistent with infinitely many different rules and principles that simply do not apply. Thus, if a language learner intends to acquire syntactic knowledge from the input, she will be more likely to get things wrong. Now, if syntax is not (for the most part) acquired from the input, language learners must already have it before being exposed to the input. The step from having enough syntactic knowledge for linguistic competence, prior to facing any linguistic input, to having a universal grammar is rather short, for such prior syntactic knowledge must apply to any possible human language.

Thus, the dispute between generativists and socio-pragmatists about syntax acquisition concerns the very nature of linguistic input and whether it is impoverished, as generativists claim, or rather rich enough to account for syntactic knowledge. This dispute, in turn, becomes a dispute concerning the nature of the cognitive abilities needed for acquiring syntax. If the input is rich enough, these abilities will most certainly *not* be a set of highly abstract grammatical principles, since a rich enough input could be comprehended by means of statistical analysis. This would go against the closed view's commitment to the existence of a language specific domain of cognition (see *cognition* above; see section 3.1).

Furthermore, if as a matter of fact linguistic input were essentially enriched by social interaction, then syntax acquisition would seem to involve the interaction between domain general cognitive abilities (e.g., statistical analysis) and socio-pragmatic abilities (e.g., intention understanding). This interaction between multiple domains of cognition would speak against the closed view's commitment to there being non-interactive, autonomous subdomains of linguistic knowledge (see *autonomy* above; see section 3.1)

So let us focus on the arguments and evidence concerning linguistic input. There are three substantial features of linguistic input that are commonly

presented as evidence of its impoverished status as a source of syntactic knowledge. First, it is commonly argued that the input is rather limited as a source of evidence, as it presents data that is logically consistent with billions of alternative ungrammatical constructions. Since it does not explicitly signal which among these alternative constructions are to be learned, linguistic input offers no clear evidence from which to acquire syntactic knowledge (see Hyams 1986). Second, it is argued that there is no negative feedback that can explain why it is that infants do not learn ungrammatical utterances. Early empirical studies on language acquisition show that linguistic input includes little negative feedback on syntactic errors (see Brown and Hanlon 1970). Further research argues that the kind of feedback that exists is too weak to support syntactic knowledge, while the needed feedback—i.e., for all errors all through development—is simply nonexistent (Marcus 1993). Third, generativists commonly point out that, even though linguistic input is so unclear as a source of evidence, and so lacking in negative feedback, young language learners successfully avoid syntactic errors that should be expected on the basis of such poor evidential set (see Bloom 1990; Pinker 1994).

To explain how it is that all human infants successfully acquire their native language, and how easily they avoid syntactic errors, even though linguistic input is as impoverished as described, generativists postulate the existence of an innate universal generative grammar (UG). This UG is a cognitive mechanism (see Ambridge and Lieven 2011) made up of knowledge of syntactic categories and rules of combination, highly abstract grammatical principles applying to syntactic categories across languages (e.g., binding, control, and structure dependence), and a sensitivity for distinct parameters that are equally abstract yet vary across languages (e.g., head-direction, null-subject).

The idea that newborns are equipped with UG, which functions as a language acquisition device (LAD), does offer support for the cognitive commitments of the closed view. The UG postulated by generativists as constituting the initial-state of language acquisition is meant to be a language-specific cognitive mechanism that accounts for linguistic competence. So its mature form seems tailor-made to satisfy the demands of *cognition*. At the same time, the principles and rules constituting the mechanism are meant to be highly abstract and purely syntactic. As such, it can be said to account for the autonomy of the syntactic subdomain of linguistic knowledge required by *autonomy*.

Unfortunately for the closed view, contemporary research shows that the initial-state of language acquisition is far from involving any domain specific and highly abstract linguistic knowledge such as the UG. Even more, there is substantial debate as to whether adult grammar does in fact constitute such a highly abstract and generative knowledge including general rules such as the principles (e.g., binding, control, and structure dependence) and parameters

(e.g., head-direction, null-subject, and specifier-head) described by generativists (see Chomsky 1981; Hyams 1986).

Recent studies have shown that there is substantial social support for language acquisition, that young infants play an active role in the learning process, and that from early on infants have an impressive understanding of the social world. In other words, linguistic input is not as impoverished as initially believed (see Hirsh-Pasek, Treiman, and Schneiderman 1984); human infants are not passively waiting for others to teach them how to engage in social interaction, rather they are actively interpreting and analyzing adults as intentionally using language (see Tomasello 2003 and 2004); and these social skills and support have an important bearing on syntax-learning process (see Hirsh-Pasek, Golinkoff, and Naigles 1996; Hoff 2006).

Our current understanding of how syntactic knowledge is acquired varies not only with respect to the nature of linguistic input, but also with respect to the role of the language learner. The poverty of stimulus argument assumes that language, as it is used by competent speakers in the environment of the language learner, massively underdetermines what the latter must know, both because it is consistent with ungrammatical structures and because it includes no corrections that can help the learner avoid such mistakes. The idea that linguistic input is not just bare language use but a rich social environment already offers a counterbalance to the poverty of stimulus argument. With enough social interaction—e.g., by communicating intentions while using language—logically consistent yet ungrammatical alternatives can be avoided. If, furthermore, the language learner is actively searching for this kind of support, then it is not surprising that systematic ungrammatical mistakes are missing. A socially competent, active language learner facing a socially enriched use of language can explain how syntactic knowledge is acquired without presupposing a UG, or so it is argued by the socio-pragmatic account.

How exactly could this take place? The socio-pragmatic account of syntax acquisition requires a radical shift from the generativist account with respect to how we understand both linguistic input as a source of knowledge and adult grammar as the end-state of syntactic development. First, linguistic input is seen, as I just said, as a rich context conducive to *intentional* social interaction. The view also appeals to an active child equipped with enough domain-general cognitive abilities to carefully study the input, including an understanding of intentions, statistical learning, pattern recognition, and structure mapping (see Saffran, Aslin, and Newport 1996; structure mapping, Genter, Holyoak, and Kokinov 2001).

Second, adult language use is no longer seen as a reflection of adult knowledge of highly abstract and general syntactic principles and parameters. Rather, adult *grammar* somehow reflects adult *language use* by regularizing

its repetitive patterns. On the socio-pragmatic account there is no universally shared syntactic knowledge across languages, what is universal is the tendency to regularize the patterns of a given socio-linguistic communicative practice. What adult speakers possess is just an inventory of linguistic expressions, which may vary according to distinct socio-linguistic environments.

Grammatical knowledge is thus constituted by a number of alternative constructions at different levels of abstraction (see Bybee 1985; Givón 1993; Langacker 1987; and Van Valin). With a big enough inventory of constructions it is possible to find common syntactic structures relative to a given language. If, as the evidence shows, from early on infants understand that adults intentionally use certain constructions for communicative purposes, it is not surprising that, as generativists claim, young infants successfully manage to avoid grammatical errors. They need only identify the constructions associated to the relevant intentional behavior.

Hence, the question of how is syntactic knowledge possible turns into the question of how is a human infant capable of acquiring the relevant inventory of linguistic expressions. To do so the language learner must find a way to identify the repetitive patterns of language use that are present in social interactive communicative contexts. Children observe adult use and study it, by focusing on their communicative intention, in order to imitate it. This delivers a pragmatic entry into syntactic knowledge by letting language learners identify the expressions that play a communicative or social-interactive role. Eventually, this will help the child form syntactic categories. Yet, imitation is not all there is to child language use, as children are capable of going beyond what they have been exposed to. Children do not only observe and imitate, they also engage in statistical analysis, pattern recognition, and structure mapping. This helps them make predictions about uses of language they have never encountered yet. Together, these cognitive abilities explain why children exhibit grammatical understanding before they can produce linguistic expressions. Adults typically use language grammatically, and their use typically exhibits the relevant patterns needed for the child to acquire syntactic knowledge.

The evidence supporting the socio-pragmatic account keeps growing. There is evidence that at around 18 months of age, children already exhibit some grammatical knowledge (see Bloom 1970; Brown 1973). The evidence strongly suggests that human infants use statistical learning to acquire it. Studies show that from early on human infants use statistical analysis to identify repetitive patterns in speech and form word-like representations based on these cues. In a study with 8-month-olds Saffran (2001a) shows that even young infants can learn to distinguish between words and nonsensical strings of sounds by means of statistical analysis of repetitive patterns in their native language. In a parallel study with infants and adults Saffran (2001b), shows

that human language learners exploit the existence of predictive dependencies—i.e., when words of different syntactic categories typically co-occur in the same sentence, such that there is a predictive relationship between one category and the other—to acquire knowledge of syntactic structure through statistical analysis.

Different studies have shown how there can be semantic bootstrapping for syntax acquisition by means of domain-general learning mechanisms that have been found in children and adults. Language learners may use their knowledge of the meaning of words to determine their syntactic category. This can be done by statistical analysis of distributional information. In a study of child directed speech Redington, Chater, and Finch (1998) demonstrate that distributional information is a formidable cue for identifying grammatical categories, thus constituting a key source of syntactic knowledge. Mintz, Newport, and Bever (2002) show that if words are categorized by their patterns of co-ocurrence with surrounding words, a distributional analysis correctly categorizes most nouns and verbs of a child's lexicon. Findings such as this strongly support the claim that children could determine grammatical categories merely by attending to repetitive patterns of word use. Other studies show that using distributional analysis does in fact constitute a strategy used by speakers in general. In a study with adult speakers facing sentences from an artificial language, Mintz (2002) shows that adults recognize words in newly presented sentences by using a distributional analysis to categorize expressions in previously encountered sentences.

In a series of studies Saffran (see Saffran 2001a; 2001b; 2002; and 2003) presents evidence suggesting that the grammatical features shared by (almost) all natural languages are themselves shaped by the constraints of the human cognitive mechanisms responsible for their acquisition. Humans are better at learning sequential stimuli if it exhibits patterns of co-occurrence that yield predictive relationships (Saffran 2002). Statistical learning, in particular, works under such constraints; it aims at identifying patterns that can yield predictive relationships among elements of the sequential stimuli. This is true irrespective of the linguistic nature of the stimuli, as it is equally demonstrated with visual stimuli. The evidence suggests (see Saffran 2003) that it is in virtue of the constraints that shape human learning mechanisms that human languages have a grammatical structure that exhibits predictive relationships among expressions in different categories. This supports the socio-pragmatic shift. It is *not* that almost all human languages present the same grammatical features because they reflect a common universal grammar. Rather, there are nearly universal grammatical features for human languages because they must all obey the same constraints for the learning mechanisms involved in language acquisition. Human cognitive limitations have shaped natural

languages into having predictive dependencies that determine their grammatical structure.

Other studies have shown that, far from being a poor stimulus, linguistic input is rich enough for human infants to learn a language, especially thanks to the social interaction that adult language use is imbued with. Young children are remarkably reluctant to go beyond what has been provided by the input (Tomasello 1992 and 2000). The evidence shows that children do not make use of their syntactic knowledge as if they possessed a fully general and abstract grammatical knowledge (e.g., a universal grammar), but do so only based on the expressions they have been exposed to. Based on this evidence, Tomasello argues that verbs are learned piecemeal, thanks to the support of social interaction and social cognitive abilities such as Theory of Mind understanding. Thus, on this view, verbs are not learned as a whole category, as it should be if infant grammatical competence were based on an innate universal grammar. There is dispute, however, as to whether it is only social interactions that determine language development or if, alternatively, linguistic and social competence co-determine each other (see Shatz 1992).

Finally, no strong empirical support seems to be available for the nativist argument. The central claim of generativist views is that children have syntactic knowledge without a corresponding experience. However, there is no clear empirical case to be made for this claim (see Sampson 2002). Lidz, Waxman, and Freedman (2003) claim to have evidence supporting the poverty of stimulus argument. They show that 18-month-old children are capable of understanding the use of anaphoric *one*, even though uses of it are rare in the input. However, Akhtar, Callanan, Pullum, and Scholz (2004) have shown that uses of anaphoric *one* are not as rare as expected and that children may end up learning it through pragmatic inferences. The dispute is open (see Lidz and Waxman 2004; Tomasello 2004), yet the poverty of stimulus claim is still wanting of supportive evidence.

To sum up, there are two alternative accounts of how infants acquire syntactic knowledge. On the generativist view, humans are innately equipped with a universal grammar that develops mainly by maturation. Children do not acquire nor do they improve their knowledge of syntax by learning from the input. On the socio-pragmatic account, human infants are not equipped with such a specialized knowledge, but rather with a substantial understanding of intentionality. They actively engage in social interaction and, through it, patiently study adult language use. Thus far, there is no clear evidence supporting the poverty of stimulus argument needed to defend the generativist account. There is, however, mounting evidence supporting the multiple aspects of the socio-pragmatic account of syntax acquisition.

There are multiple studies showing that children are no passive language learners, exhibiting an understanding of intentionality from early on (Saffran 2001a; Tomasello 1992 2000; Shatz 1994). Linguistic input always takes place in social-interactive settings, with enough support from adults to aid the learning child (Saxton, Backley, and Gallaway 2005; Chouinard and Clark 2003). Statistical analysis of distributional information of the input seems to get grammatical categories right (Mintz, Newport, and Bever 2002). In general, human subjects are better at learning sequences of stimuli if they exhibit repetitive patterns (Saffran 2002). Specifically, humans are better at learning *languages* with repetitive patterns (Saffran 2001b; 2003). And finally, both infants and adults, do use statistical analysis and pattern recognition as learning mechanism (Saffran, Aslin, and Newport 1996; Mintz 2002).

The evidence strongly suggests that the generativist view is unsupported. There is no need to postulate an adult generative grammar. A rich input in an interactive context, with adults sharing syntactic knowledge through social interaction (Bruner 1975; Halliday 1975) is needed. With enough time, this active social child, aided by her impressive domain general analytical capacities, will have acquired a big enough set of alternative constructions that will become an adult grammar once regularized (Bybee 1985; Givón 1993; Langacker 1987; Van Valin 1993).

The evidence from syntax acquisition goes against the cognitive commitments of *strong compositionality* (see section 3.1). Against *cognition*, the evidence shows there is no language specific cognitive mechanism involved in the acquisition of syntactic or grammatical knowledge. Syntactic knowledge appears to be acquired through the interaction of multiple domain-general mechanisms, such as statistical learning from distributional information, pattern recognition, and intention understanding, all of which benefit from knowledge in other domains such as Theory of Mind. This goes against *autonomy*. Far from being modular, impenetrable, and autonomous, syntactic competence appears to be a highly interactive, domain general, and penetrable cognitive capacity. It is important to remember that these conclusions do not only pertain to infant language use. The socio-pragmatic account is concerned not only with the initial stages of language learning but also with the end state, which is understood as knowledge of a set of regularized, alternative grammatical constructions.

It is important to note that some developments within generativist accounts may be seen as offering substantial objections against socio-pragmatic explanations. Halle and Marantz (1993) forcefully argue that there is no substantial syntactic distinction between words and sentences, as they are all product of the same procedures. The view, known as Distributed Morphology—see also Halle and Marantz 1994; Embick and Marantz; 2008—claims to have a

satisfactory account of certain linguistic phenomena that the socio-pragmatic accounts fails to explain. Determining the value and strength of these objections is beyond the limits of this book. Nonetheless, even if accepted, these objections are no good news for the *closed view*, since distributed morphology is not compatible with *strong compositionality*. According to distributed morphology, syntactic compositions in the form of morphological constructions permeate all through natural languages, including words and not only sentences. Words themselves are the result of morphological composition. Thus, on this view, there is no basic *lexicon* from which sentences are built, and thus there is no intrinsic meaning that words have in isolation and, consequently, no autonomy of the lexicon with respect to the syntax. Thus, the idea of a specific domain of lexical, syntactic, and semantic knowledge following a unidirectional, bottom-up procedure of word-based composition of complex meanings—as *strong compositionality* requires—makes little sense on this view.

Still, there is one possible reply that, if successful, may stir the *closed view* out of the empirical trouble set up by its cognitive commitments. According to this reply, the principle of *strong compositionality* must (for some reason yet to be offered) not be considered as a claim about human cognitive abilities in any way shape or form. It must be understood as offering an account of how competent speakers go on interpreting complex expressions. On this view *strong compositionality* is to be strictly understood as a claim about natural language processing (see section 3.2). In the following section I will consider this alternative interpretation of the closed view and its principle of *strong compositionality* and present empirical studies addressing the issue. As I will show, the data demonstrate that natural language processing by competent speakers simply does not verify *strong compositionality* or its operational version *processing*.

3.6 LANGUAGE PROCESSING

The empirical evidence presented in sections 3.4 and 3.5 strongly suggests that *strong compositionality* fails to offer an account of what speakers know when they know a language and how they manage to acquire and sustain such knowledge. Alternatively, *strong compositionality* may be interpreted as offering a processing account of language that is not concerned with its cognitive architecture. On this interpretation *strong compositionality* is not about speaker knowledge of language. After all, *strong compositionality* is about meaning determination for complex expressions in a representational system. Whether or not speakers *know* this is irrelevant, or so it is claimed. Natural

languages may still be strongly compositionally so long as the meaning of all complex expressions in natural languages is compositionally determined.

Before considering this alternative interpretation (see section 3.2), it is important to keep in mind how problematic the view is. If this reading is to avoid the empirical challenges presented in sections 3.4 and 3.5 it must claim that natural languages may be strongly compositional even if speakers, competent speakers that is, do not happen to know that they are, or even possess *any knowledge* compatible with it. This possibility seems rather strange. It is possible, of course, for competent speakers to fail to have *reflective* knowledge of the compositional nature of the language they happen to master. But it seems far-fetched to think that competent speakers may also fail to possess, even at a sub-personal level, any knowledge even compatible with such compositional nature. If such were the case, then it is utterly mysterious how it is that competent speakers use and understand a strongly compositional language. It cannot be because they *know* how to do it, or that they have a dedicated cognitive structure to do it. They just, somehow, do it. Briefly put, to claim that *strong compositionality* is merely a processing principle with no cognitive commitments attached is tantamount to claiming that there is no account of the competence of competent speakers. Unless, of course, one wants to claim that speakers do so solely by means of their general knowledge and reflective abilities, in pretty much the same way a theorist has to do to understand how language works. On this view, competent speakers process languages compositionally because they want to do so, or simply because. Even before considering the empirical evidence, this "*pure processing*" interpretation of *strong compositionality* is highly problematic, to say the least. However, I will ignore this in what follows. As I will show, the empirical evidence suggests that *strong compositionality* is false even when understood merely as a processing account.

According to the processing interpretation of *strong compositionality* the meaning of all complex expressions of a natural language is determined in accordance to *processing*.

Processing: For any competent speaker *S* and any complex expression *E* of any natural language £, *S* understands the meaning of *E* in £ if and only if *S* processes *E* by following a bottom-up direction of algorithmic calculation, without any backwards (top-down) steps of re-calculation prior to the conclusion of the algorithmic procedure.

As we saw in section 3.2, *strong compositionality* is committed to the claim that competent speakers follow a finite, unidirectional, bottom-up procedure to determine the meaning of any complex expression in their

language. Speakers, on this view, (i) identify and process expression parts and their meanings; (ii) identify the relevant syntax; (iii) and deliver and output after applying the corresponding semantic rules. The process is unidirectional and bottom-up since it goes from the meanings of expression parts to the meaning of the complex expression (i.e., bottom-up) and it is never allowed to go backwards (i.e., unidirectional). Allowing for backwards steps (i.e., top-down) would go against the logical closure of the compositional process by either rejecting the autonomy of syntax, by revising the corresponding syntactic structure once we have the meaning of the complex expression; rejecting the autonomy of the lexicon, by revising the meanings of expression parts based on syntactic considerations, or both, by using pragmatic considerations to reinterpret both the syntactic structure and meanings of expression parts.

As I said in 3.2, to properly understand *processing* we must look at the logical nature of the proposal, as it is meant to describe a *closed* procedure. As we saw in chapter 2, two features are essential to the *closed* nature of the linguistic domain according to *strong compositionality*, namely, its impenetrability *qua* domain and the autonomy of its subdomains. These same features must take stage when it comes to processing. If the linguistic domain is, in terms of language processing, impenetrable by information from external domains, then the output delivered after following the bottom-up procedure must be complete; it cannot lack any meaning and/or truth-conditions. If there were any missing parts, then the interpretation process would not have concluded, it would still be open, allowing for backwards steps of revision based on information from other domains. This does not mean that *processing* does not allow for any kind of, say, pragmatic revision or reinterpretation. Such modification is allowed, but only after the algorithmic process of compositional interpretation has concluded. In other words, only after the compositional process has delivered a complete, meaningful, and truth-evaluable interpretation is it possible to engage in reinterpretations. Furthermore, such revisions must be motivated only by non-linguistic considerations and not by any lexical, semantic, or syntactic need since no such interpretation can alter elements of the algorithm by including non-linguistic information. Similarly, if the subdomains of linguistic processing—lexical, syntactic, and semantic—are to be autonomous from each other, there cannot be any influences or interventions amongst each other during the interpretation process. There is no backward processing, even within the linguistic domain. Each step in the calculation must go forward until a complete assignment has been given. No step can move backwards and revise the results obtained in any prior step. There are no semantic revisions of syntax and no syntactic revisions of the lexicon. Competent speakers simply identify the lexicon and the syntax, and then apply the relevant semantic rules to the lexical meanings in accordance

to the syntactic structure. Non-linguistic revisions appear only after a complete meaning assignment has been given.

Unfortunately for *strong compositionality*, the empirical evidence on language processing falsifies this view. Garden-path sentences, sentences exhibiting syntactic or semantic ambiguity, provide an excellent case study for determining the structure of language processing. A proper interpretation of a garden-path sentence *requires* the use of non-linguistic information to solve the ambiguity. If *processing* is correct, then such information must be used only after the compositional process has concluded and delivered a complete interpretation.

Consider the case of garden-path sentences exhibiting referential ambiguity. There is referential ambiguity whenever a speaker uses a referential expression in a context where two or more objects are equally good candidates for the referent of the expression—e.g., when a speaker uses the expression "The martini glass" in a context that includes two or more martini glasses. According to *processing* the hearer should construct a complete representation of the meaning of "The martini glass" by merely using lexical, syntactic, and semantic information, and in that order. Pragmatic information involving contextual considerations—such as the interests of the speaker or even the visual perspective of the speaker—should be taken into consideration only if the ambiguity persists after a complete meaning has been determined. If non-linguistic information is taken into consideration prior to the conclusion of the compositional procedure then *processing* and, with it, *strong compositionality* would turn out to be empirically false.

Recent studies have demonstrated that competent speakers, infants and adults, do take into consideration non-linguistic information in the early stages of language processing and, hence, prior to completion of what should be an algorithmic compositional procedure. Nadig and Sedivy (2002) studied how 5 to 6 year olds solve referential ambiguity tasks. To complete the task children had to follow instructions from an adult confederate and pick up a target object from a visual display. The display included four objects, one of which was occluded from the adult. The adult turned around to avoid seeing the objects as they were placed on the display and then faced the child when giving instructions on how to move the objects. Critical instructions included expressions such as "Pick up the glass." Two conditions were compared. In the "common ground" condition, there were two glasses (a big one and a small one) visible to both the adult and the child. In the "privileged ground" condition only one of the glasses was visible to the adult and the other one was visible only to the child. The goal of the experiment was to determine, first, if 5 to 6 year olds make use of pragmatic information—i.e., information about other's perspective and about a shared perspective between speaker and hearer—when processing language; and, second, if they do so, the

experiment aimed at determining when exactly, at which stage of language processing, is this information taken into account.

To determine whether and when did children make use of extra or non-linguistic information to interpret the expressions, the experimenters used a head mount eye tracker on the experimental subjects. After reviewing the video recordings, the experimenters were able to determine exactly which objects were being looked at by the subjects and at which precise moment during the task. The experimenters expected children to take longer to select the target object in the common ground condition—i.e., where two referential candidates were available for both the child and the adult—compared to the privileged ground condition—i.e., where only one referential candidate is available for both participants. They also expected children to consider both objects as equally good referential candidates in the common ground condition, thus exhibiting a corresponding eye movement pattern going from one to the other object. If the evidence confirmed these expectations, it would demonstrate that 5 to 6 year olds do take extra-linguistic (i.e., pragmatic and common ground) information into account when parsing language. Now, if in fact both conditions differed with respect to eye movement patterns, determining when exactly does this difference take place would help determine whether pragmatic information is taken into account *prior* to, or *after*, the completion of linguistic processing. If pragmatic extra-linguistic information constraints the interpretation *from early on* then the eye movement pattern of each condition should differ as soon as the noun "glass" is presented in the instruction "Pick up the glass."

The results form the experiment confirmed both expectations. First, children as young as 5 years of age "use common ground information with striking speed and efficacy to constrain temporary indeterminacies." (Nadig and Sedivy 2002, 334) Second, and most important for our discussion concerning the empirical adequacy of *processing*, the use of extra-linguistic information appears *from early on* in the interpretation, *prior* to the completion of what should be the compositionally closed, bottom-up interpretation procedure. The results show that "by the offset of the noun (approximately 560 ms after the onset), fixations in the privileged ground condition and the common ground condition began to differ, indicating that fixations programmed prior to the end of that word were influenced by common ground information." (Ibidem) Briefly put, 5 to 6 year olds start taking extra-linguistic information into account when interpreting an ambiguous linguistic expression even before the speaker has finished uttering it. This is evidence of pragmatic, extra-linguistic information being part of the meaning determination process *prior* to its compositional conclusion, thus falsifying *processing*.

Hanna, Tanenhaus, and Trueswell (2003) designed a very similar experiment this time involving adult speakers. Subjects (undergraduate students)

were asked to solve a referential ambiguity task in both common ground and privileged ground conditions, in line with the Nadig and Sedivy (2002) experiment. Hanna and colleagues added two new elements. First, an experimenter described the subject's display of objects to the speaker, a confederate participant in the task, who is meant to give instructions to the subject (addressee). The display was described either accurately or inaccurately in a way that only the subject (addressee) is aware of. This generated a match or a mismatch between the subject's display and the speaker's representation of it. This made it difficult for the subject to take into account the speaker's perspective, which corresponds to the inaccurate description offered by the experimenter early on, forcing her to make use of her short term memory. Second, Hanna and colleagues used instructions involving referentially ambiguous expressions that could disambiguate earlier or later in the instruction depending on the situation. For example, there was early disambiguation if the subject was told to "Pick up the empty martini glass" when the display included two empty jars and one empty martini glass. As soon as the subject hears the word "empty" she already knows that the martini glass is the target. The same instruction, "Pick up the empty martini glass," would have late disambiguation if the display included two pairs of empty containers, two jars and two martini glasses. It is not until she hears the word "martini" that the subject can identify the empty martini glass as the target object.

Like Nadig and Sedivy (2002) Hanna, Tanenhaus, and Trueswell (2003) asked, among other things, *whether* and, if so, *when* do subjects make use of information concerning the speaker's perspective to resolve the referential ambiguity carried by the speaker's instruction. Addressees achieved referential resolution earlier (about 300 milliseconds) in cases of early disambiguation than in cases of late disambiguation. Given the instruction "Pick up the empty martini glass," addressees fixated on the target object about 600 milliseconds after the onset of the determiner "the" in early disambiguation contexts, and 900 milliseconds after in late disambiguation contexts. These results show that "addressees were rapidly able to take the speaker's perspective into account" (Hanna, Tanenhaus, and Trueswell 2003, 57). Consistent with the results obtained by Sedivy, Tanenhaus, Chambers, and Carlson (1999), subjects appear to be performing an incremental processing of syntactic and semantic interpretation paired with an early use of pragmatic and, in this case, perspectival information even before the instruction as a whole has been parsed. Interestingly, the integration of pragmatic information prior to conclusion is confirmed even for mismatch scenarios, even though in those cases subjects had the extra burden of using her short term memory to recall what the experimenter said in order to figure out the speaker's perspective, given that it conflicts with what is perceptually available to her.

These studies confirm that competent speakers, both children and adult, integrate non-linguistic information—beyond the lexical, syntactic, and semantic information considered by *strong compositionality*—from early on in the interpretation process. This contradicts *processing*. Further studies offer more evidence against this principle by showing that the process rarely delivers a complete representation and, hence, that the compositional procedure rarely reaches its alleged conclusion.

According to *processing* (see above) competent speakers understand any given natural language sentence by constructing a *complete representation* of its meaning following a bottom-up procedure, from lexical meanings to semantic composition, all guided by a detailed understanding of the sentence's syntactic structure. The only alternative to constructing a complete representation as a result of the interpretation process is to offer no representation, but this result is limited to cases of ungrammaticality. This is a consequence of the logically closed character of strong compositionality. If, given a syntactically and semantically complete sentence *E*, the compositional procedure delivers an incomplete representation *R* then it follows that the semantic and syntactic elements of *E do not* completely determine its meaning. If any such thing were to take place, it would be tantamount to a rejection of *strong compositionality*.

Ferreira, Bailey, and Ferraro (2002) and Ferreira and Patson (2007) review a series of studies concerning how adult speakers understand sentences. The findings are outstanding both for their breath and substance. There are multiple data showing that, in general, competent speakers have a very shallow understanding of sentences, nonetheless this understanding is enough for their communicative purposes. Far from having a complete and detailed representation of a sentence's meaning, the evidence shows that speakers construct incomplete representations, in some cases this results in misunderstanding *the allegedly complete meaning* of the relevant sentences, even non-ambiguous ones.

Cases of semantic illusion and garden path sentences support these claims. Semantic illusions appear when subjects systematically understand sentences as having a certain meaning that explicitly differs from its literal one. The Moses and Survivor illusions are well known. When asked how many animals of the same species did Moses bring to the ark, people systematically reply by saying "2," ignoring that, according to the biblical story, it was Noah, not Moses, who did it (see Erikson and Matteson 1981). Similarly, competent subjects typically consider the question "where should the authorities bury the survivors?" as an acceptable one, ignoring the abnormality in burying people alive (see Barton and Sanford 1993). Sanford and Sturt (2002) look at these and further studies to argue that underspecified interpretations play a central role in actual human understanding of language. Aside from semantic

illusions, there is further evidence that, contrary to what *strong compositionality* claims, word meaning is not fully employed in linguistic interpretation and that non-linguistic information and general knowledge become part of the interpretation process from an early stage. For example, the sentence "No head injury is too trivial to be ignored" is systematically interpreted as saying "No matter how trivial it might appear to be, a head injury should be treated." Interestingly, this latter, correct interpretation is ruled out by any compositionally closed interpretation—its syntax does not correspond to that of the original sentence—and, furthermore, an alternative paraphrase of the same original sentence is equally justified yet mistaken, namely, "No matter how trivial it might appear to be, a head injury should be ignored" (see Wason and Reich 1979; and Natsopoulos 1985).

Other studies show that adult speakers make use of general heuristic principles, and not only syntactic algorithms, in early stages of sentence processing. In a study with undergraduates Ferreira (2003) found out that almost 30 percent of subjects misinterpret ordinary sentences if they appear in passive form. They mistakenly judged that "dog" was the agent of "The dog was bitten by the man." This proportion of misinterpretation is surprisingly large given that 99 percent of subjects were accurate when interpreting the same sentence in active form. This difference, however, did not arise when considering syntactically complex versions of the same sentences (i.e., It was the man who bit the dog). These results suggest that the surprising misinterpretation is not owed to either syntactic complexity or a lack of frequency of syntactic form. To account for the difference between active and passive forms Ferreira (2003) argues that speakers use simple heuristic principles—e.g., the first noun phrase is the agent the next noun phrase is the recipient of the action—to interpret English sentences. Heuristic principles, such as the *first noun agent* principle, take place in language processing just as much (if not more) as algorithmic syntactic processing. Van Herten, Kolk, and Chwilla (2005) studied how Dutch speakers processed semantically implausible sentences—e.g., stating that a fox hunted a poacher. Van Herten and colleagues measured the brain waves of 42 adult Dutch speakers and discovered that implausible sentences elicited a P600 event related potential (ERP), associated with syntactic processing. Van Herten and colleagues conclude "participants used a plausibility heuristic that made them assume the reading of the sentence that fits most with their world knowledge." (Van Herten, Kolk, and Chwilla 2005, 254)

Finally, multiple studies of brain activity during language parsing using ERP measurements have demonstrated that sentence processing is much more than just following the bottom-up procedure of syntactic and semantic combination of lexical meanings (see van Berkum 2008 for a general review). Competent speakers do not even wait to hear the end of a word, let alone the

end of a sentence, to rapidly use all sorts of information in order to arrive at the earliest possible interpretation. When interpreting language, speakers anticipate, by means of both sophisticated statistical analysis and simple heuristic reasoning (see Otten and van Berkum 2007; Nieuwland and van Berkum 2005). Speakers anticipate discourse to arrive at the earliest possible interpretation, "and if it makes sense, they sometimes ignore the syntactic rules of their language. What we see is an opportunistic, proactive brain at work." (Van Berkum 2008, 379)

The studies show that subjects rapidly take into account multiple sources of information—information about the speaker, general knowledge about the world, and contextual information about the wider discourse—from early on in sentence processing. Subjects were found to fix on an interpretation of a word depending on context coherence even before they hear the end of the word (Camblin, Gordon, and Swaab 2007). Subjects also appear to draw information about the identity and perspective of the speaker from early on when interpreting a word (van Berkum, van den Brink, Tesink, Kos, and Hagoort 2008). Referential resolution constitutes a special case in language processing, as the need to resolve referential ambiguities appears to have a special pull on subjects, who feel forced to find an adequate referent even at the cost of revising the syntax of the corresponding sentence (see Kuperberg 2007; van Berkum, Koornneef, Otten, and Nieuwland 2007; van Berkum, Brown, and Hagoort 1999). Simply put, "referential factors can sometimes briefly lure people into pursuing a syntactic analysis that is ungrammatical [. . .] [W]hen interpreting language, people don't just slavishly follow syntax." (van Berkum 2008, 378) Thus, contrary to what *processing* states, there is non-linguistic information playing a role from early on in sentence processing and, furthermore, there is evidence of backward steps of reinterpretation that aim at revising the syntax of the sentence to disambiguate the referent of a term.

Based on these empirical data one can convincingly argue (following Ferreira, Baily, and Ferraro 2002; van Berkum 2008; and Jackendoff 2007 among others) that the architecture of language processing is heavily constrained by contextually determined goals. Typically these goals include, for example, offering a proper response or follow up in a dialogue, nodding approvingly, or engaging in some non-linguistic motor action. Very rarely do these goals require that the hearer form a syntactically complete and accurate representation of the speaker's utterance. Instead, speakers form representations that are simply *good enough* to achieve the relevant goal. Such representations, however, do not presuppose completing an algorithmic compositional procedure; whatever gets speakers to the earliest interpretation will do. Sometimes simple heuristics and even general knowledge will do, sometimes more sophisticated statistical analysis will be needed. In fact,

following a strict compositional procedure will sometimes deliver an incorrect interpretation of the corresponding sentence, an interpretation that fails to meet the participant's communicative goal(s) in the conversation.

Given its empirical failure as a cognitive hypothesis about human linguistic competence (sections 3.3 to 3.6), in this section I have considered the *processing* interpretation of *strong compositionality*. On this alternative understanding, *strong compositionality* is to be viewed as an account of how competent speakers interpret complex expressions. However, the evidence on actual language processing demonstrably discards the resulting *processing* principle and, with it, *strong compositionality's* view of natural language understanding. This may seem surprising since a *compositionally closed* processing of sentences seems *prima facie* the most intuitive or naïve one: "[f]irst you recognize each of the words, then you look up their meaning in your mental dictionary, and then, using syntax to guide the combination, you simply combine the meanings so that you know what is said." (van Berkum 2008, 376) Still, the evidence shows that competent speakers do make use of non-linguistic information from multiple sources, and they do so from early on in the process, before they even hear the end of the complex expression and, thus, clearly before any algorithmic syntactic processing may reach its conclusion. Even more, competent speakers do engage in top-down and backward steps of revision (both semantic and syntactic), in order to achieve a proper interpretation. Far from being bottom-up, forward-looking, and algorithmic, actual sentence parsing appears to be multidirectional, making use of information from multiple sources, and guiding itself by fast and frugal heuristic reasoning.

3.7 NATURAL LANGUAGES AND STRONG COMPOSITIONALITY

If *strong compositionality* is meant to be a hypothesis (or a methodology) about natural languages, then its truth or adequacy depends on whether humans understand natural languages according to it. To find this out we need to operationalize *strong compositionality* based on its psychological assumptions concerning, first, what speakers must know in order to learn and exercise a human language (i.e., cognitive assumptions) and, second, what speakers must do in order to properly understand a complex expression of a given natural language (i.e., processing assumptions). With operational versions of *strong compositionality* in hand, I have presented dozens of empirical studies, all of which strongly suggest that each and every one of the operational principles is empirically false.

Sections 3.3 to 3.5 develop the empirical argument against *strong compositionality* as a cognitive hypothesis. The evidence shows that humans are

not equipped with a cognitive architecture that could possibly correspond to the acquisition or exercise of a compositionally closed system of representation. Natural languages are the product of the interaction between multiple domain general mechanisms. There is no language specific domain of cognition, module or mechanism. As Shatz (2007) puts it, "currently, the "language acquisition device" can be described as a probabilistic thinker, capable of making inductive inferences across utterances over time based on a broad array of cues changing with development." (Shatz 2007a, 9) The studies show as well that this interaction also takes place at the subdomain level. There is substantial evidence of syntactic bootstrapping for word learning as well as semantic bootstrapping for syntax.

Section 3.6 develops the case against *strong compositionality* as a processing hypothesis. Multiple studies on actual language processing show that speakers make use of non-linguistic information from early on in the interpretation process; that they make top-down syntactic and semantic revisions, sometimes even considering ungrammatical interpretations to resolve ambiguities; that the interpretation does not always result in a complete and syntactically detailed representation; and that interpreters commonly anticipate the words of the speaker in order to form a representation even before they can identify the target sentence's syntax. Thus, competent speakers do not in fact understand complex expressions by following anything like a compositional procedure, much less the one offered by *strong compositionality*. Actual sentence parsing is multidirectional, draws information from varied sources, and may even follow heuristic strategies.

Strong compositionality distorts what speakers know about natural languages and how they acquire them, and it wildly misrepresents how actual human subjects parse complex expressions. This leaves the closed view with little elbow room. If *strong compositionality* is not an account of what speakers know when they know a language, and it is also not an account of what speakers do when they use and understand a language, then it is hard to imagine what it is an account of. It cannot be an account of what competent speakers *could possibly know* even though, as a matter of empirical fact, they do not. This could only be the case if humans where to have, in such counterfactual scenario, a substantially different cognitive endowment than the one they actually possess, for example, by having a domain specific, encapsulated, language acquisition device. This counterfactual possibility, however, requires one to move beyond what is psychologically possible to human beings. In a strict and substantial sense, what is psychologically possible for human beings is determined by whichever cognitive capacities human beings actually have. In this strict and substantial sense, human beings cannot possibly acquire, develop and maintain a natural language observing strong

compositionality. Of course, there is a careless and shallow sense in which it may be truly said that *strong compositionality* accounts for what competent speakers *could possibly know*. If so one could just as easily claim that, had human beings been different, natural languages would have been different too. The empirical evidence, of course, does not go against this rather weak claim. It goes without saying that as theorists of natural languages we are concerned with what is possible for humans as they actually are, not with what is possible for humans as they could have been.

Strong compositionality cannot either be seen as an account of what competent speakers *aim at* doing when processing language even though, as a matter of empirical fact, they do not do so by following a logically closed compositional algorithm. This could only be the case if the goal of sentence processing was, in general, to come up with a syntactically complete and truth-evaluable interpretation that would correspond with the one determined by *strong compositionality*. This, however, is known to be false. The studies discussed in section 3.6 show not only that speakers do not follow an algorithmic procedure when parsing sentences; it also shows that forming a syntactically complete interpretation is not always the goal. Sometimes it is simply enough to form a partial interpretation, one that will not even be truth-evaluable. Any prediction coming from *strong compositionality* is bound to be wrong about these cases.

Furthermore, as evidenced by sentences exhibiting syntactic or semantic ambiguity, and even some non-ambiguous ones, there are many cases where the predictions of *strong compositionality* are simply off the mark. Speakers resolve the ambiguity by directly picking an adequate interpretation without first forming a (sometimes multiple) compositionally determined one and then revising it, contrary to what *strong compositionality* demands. In many cases, the adequate interpretation does not even observe the grammar or syntax of the target sentence. The "head injury" example—"No head injury is too trivial to be ignored"—is an interesting case in point. The syntactically determined interpretation—*all head injuries should be ignored*—is the exact opposite of the adequate interpretation—*No head injury should be ignored*—yet competent speakers *directly* pick the latter. Thus, even when it may be said that speakers aim at syntactically complete and truth-evaluable interpretations, in many such cases competent speakers are *not* aiming at an interpretation that may correspond to the one determined by *strong compositionality*, and they need not go through such interpretation to find the one they are aiming at.

It appears, then, that *strong compositionality* is substantially mistaken. It is not an account of what speakers *do know* about language or *how* they come to know it; nor is it an account of what speakers *could possibly* know

about natural language. *Strong compositionality* simply offers a false account of what it is to know a natural language. It is also not an account of what speakers *do in fact* when processing natural language expressions; nor is it an account of what speakers are aiming at when processing complex expressions. The view *also* offers a false description of what speakers do and of what they are aiming at doing when engaging in sentence interpretation. *Strong compositionality* does not and cannot offer a correct account of human linguistic practice.

As theorists and students of natural language, we can do better and endorse an alternative, empirically adequate explanation. In the following chapter I will present some such account.

Chapter 4

Open Compositionality and the Cognition-First Methodology

If not *closed compositionality* what then? The quick answer has already been given: *open compositionality*. Presenting and developing the long and detailed version of the answer will take this chapter. But before doing so, let me go back to the beginning of this book.

In chapter 1, I distinguished between "two views" of language. First, the closed view claims that language is a system of representation whose main goal is to communicate thoughts among users. To explain how speakers manage to use complex expressions to achieve such goal, the communicative view endorses a logically *closed* hypothesis of *strong compositionality* according to which the meaning of any complex expression is a function of the meaning of its expression parts and the way they are syntactically combined. There is successful communication whenever the truth-conditions of this compositional meaning coincide with those of the relevant thought to be communicated. Language use, on this view, is the product of a logically *closed* system that operates upon a purely linguistic set of information. Second, the open view claims that language is a special cognitive tool useful for knowledge acquisition and, more generally, cognitive development. To explain how language can be such a powerful cognitive aid, the open view claims that language is not a closed, cognitively impenetrable, domain specific mechanism or ability but rather an open multidirectional, higher order, domain general, and interactive cognitive system. Linguistic cognition, on this view is the result of the interaction between multiple domain general mechanisms such as intention understanding, Theory of Mind, statistical analysis, lexical and syntactic knowledge, and general knowledge among others.

As I said in chapter 1 (see section 1.1) by endorsing *closed compositionality*, the closed view gains an extraordinary simplicity and efficiency of methodology. It is clear what one as a theorist should do to account for a given

meaning-related phenomenon. Assuming that natural language observes *strong compositionality* immensely reduces the alternative ways in which a complex expression may be used to convey a certain meaning. We as speakers have intuitive access to the meaning of the complex expression and there are only two variables that we must deal with, namely, the meaning of the expression parts and the syntactic structure of the whole complex expression. Once we have identified one of the latter we can deduce the other. With its endorsement of *closed compositionality*, the communicative view is pretty much at its peak. Nowadays philosophers of language and linguists are not only extending compositional semantics to account for every detailed part of language (see e.g., Elbourne 2013; Chierchia 2013; Burnett 2017; Lasersohn 2017; Lassiter 2017), they are trying to defend closed compositionality even from basic ontological problems (see Podlaskowski 2018). Unfortunately, these important theoretical advantages come at an important theoretical cost, as *closed compositionality* suffers from an almost total lack of independent empirical support. Semantics, on this closed view, has traditionally worked in a psychological void, detached from any actual requirements for human cognition. It works under the methodological assumption that, as Montague famously puts it, "there is no important theoretical difference between natural languages and the artificial languages of logicians." (Montague 1970b, 222) As I argued in chapter 3, more than a hundred years of empirical studies on language acquisition and development have shown this assumption to be false. The empirical evidence demonstrates that *strong compositionality* is simply wrong about actual human speakers. Whether we take it as a cognitive hypothesis, concerning what competent speakers know about language and how they come to know it, or as a processing hypothesis, regarding how competent speakers understand natural language sentences, *strong compositionality* misrepresents its target. There is no logically closed system of cognition that may be said to correspond specifically to language, and humans do not follow any bottom-up, forward looking algorithmic procedure when interpreting each other linguistically.

The closed view is, thus, in serious trouble. It essentially relies on the assumption that *strong compositionality* is somehow correct. The empirical inadequacy of the view befalls given the blatant falsity of its central assumption. If the closed view is to be kept, and I believe it should, the assumption that natural languages observe *strong compositionality* must be abandoned. What is left of the view if we take away the mistaken assumption? In chapter 1 (see section 1.1), I presented the closed view as addressing the issues of how linguistic practice takes place; what is the relation between what speakers say, what they think, and what they do; and how does what speakers say to others relate to what others reply to them. To address these issues the closed view helped itself with two central assumptions, namely, *correspondence* and

strong compositionality. Even though there is mounting empirical evidence against the latter, the *correspondence* assumption still stands. According to it, ordinary language use somehow reflects a speaker's mental states; speakers use natural language to express their thoughts. There is, thus, a *correspondence* between the content of a speaker's thoughts and the meaning of her words or complex expressions. This correspondence between linguistic meaning and mental content is partly explained by the conventions that give place to a linguistic practice among the members of a population. Natural languages are, after all, social phenomena. Following Lewis (1975), we might say that these conventions constitute a set of regularities in action and/or belief, among the members of a population, which helps them solve the coordination problem of finding out one another's thoughts about their surroundings.

Now, one cannot simply abandon the assumption of *strong compositionality* and keep the closed view. The conventions of language use only *partly* explain the correspondence between mental content and linguistic meaning. Hearers must already be capable of determining the meaning of the complex expressions used by the speaker before the conventions can help her find out what the speaker has in mind. To be members of a linguistic population, subjects must be able to determine the meaning of any complex expression they encounter. This is where the *strong compositionality* assumption played its role in the closed view. If language is *strongly compositional*, then the ability to determine the meaning of any complex expression is easily explained, as speakers need only acquire a lexicon, syntactic knowledge and the relevant semantic rules of composition. Given that natural languages (see chapter 3) are *not strongly compositional*, we must offer an alternative explanation of the communicative success of language users. How is it that actual competent speakers determine the meaning of complex expressions they encounter? If not *strong compositionality* what then? The multiple empirical studies presented in chapter 3 give us a hint of the alternative account I want to propose. Let me begin by stating a basic commitment of this *open* view.

4.1 THE LEWISIAN COMPROMISE

While assuming that languages, all possible languages, are formal entities constituted by sets of ordered pairs of sentences and meanings, Lewis (1975) argues that *natural* languages are a proper subset of them, distinguished by the fact that they are used by a given human population while observing the convention of truthfulness and trust in the said formal entity. Lewis is supposing, of course, that the said languages are strongly compositional in that it

is possible to offer an algebraic account of the complete set of corresponding ordered pairs of sentences and meanings. There is, however, a wrinkle. As Lewis (1975) considers the first objection to this view, many sets of ordered pairs that meet the above description "are not really possible languages. They could not possibly be adopted by any human population." To avoid this problem, we must make a compromise and accept that "(a) language in a narrower and more natural sense is any of these entities that could possibly—possibly in some appropriate strict sense—be used by a human population." (Lewis 1975, 171). What could an appropriate strict sense of "possibly" be in this case? The obvious answer is the sense in which something is *cognitively* or *psychologically* possible for humans. In order for something to be a natural language, it must be *psychologically* possible for humans to learn it, develop it, and use it. I call this the "Lewisian Compromise."

Lewisian Compromise: Natural languages are, first and foremost, things that can be learned, developed, and used by human beings given the limits and nature of their cognitive resources.

According to *open compositionality*, the first step towards a satisfactory and empirically adequate account of natural languages is an endorsement of the *Lewisian Compromise*. It is a compromise because it demands a concession from those who—like Lewis himself—believe that natural languages are set-theoretical entities that can be fully described by means of formal tools. Not surprisingly, perhaps, philosophers of language, including Lewis, have done little more than paying lip service to the compromise. It is time for us to take it seriously.

One rather important consequence of endorsing the *Lewisian Compromise* is the rejection of *strong compositionality*. Natural languages are effortlessly acquired, learned, and developed by human beings, all thanks to human cognitive endowment. No special training or study is required. As such, natural languages *cannot* be *strongly compositional*. Human beings (see chapter 3) do not have the cognitive resources required for effortlessly learning or developing a *strongly compositional* system of representation. There is no especially dedicated, domain specific, impenetrable, and autonomous cognitive capacity for language. And the—interactive and multidirectional—way in which human beings actually use their cognitive endowment to understand natural languages is far from meeting the strictures of a logically *closed* algorithmic process. With no domain specific capacity and no domain general resource—or combination thereof—available, the evidence strongly suggests that human beings *do not* have the cognitive resources required for effortlessly learning, developing and using a *strongly compositional* language.

Some may choose to ignore these results and insist on offering a compositional formal semantics to describe what speakers do, e.g., determining truth conditions. I have already said (see section 3.7) why this delivers seriously misguided predictions, as speakers sometimes do not even aim at truth-evaluable interpretations. There is a more general way of understanding what is wrong with this approach. As Segal (2001) puts it, theorizing about language in this way is equivalent to "theorizing in the vacuum." I have demonstrated that merely following the compositional methodology without proper empirical restrictions delivers trivial (see sections 2.4 and 2.5) or false results (3.6). The empirical evidence shows that following *strong compositionality* in our theorizing is to greatly detach from what human beings know and what they do with what they know. This seems to me to be a serious blunder. I agree with Dummett (1973) in that

> 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, 92)

Still, we may choose to benefit from Lewis' terminology. Natural languages may, after all, be usefully described as sets of ordered pairs of sentences—or, more generally, strings of sounds and marks—and meanings. We just need to add that the set is not *compositionally closed*; in other words, there is no mathematical function (or set of functions) determining which meaning corresponds to which sentence for every possible sentence of the relevant language. Some sentence-meaning pairs may in fact be the result of a syntactically guided composition of lexical meanings, but many others will instead be the result of early processing of non-linguistic information such as general knowledge, the intentions of others, or even result from fast heuristic reasoning. Since it is not possible to predetermine which meanings, or even which meaning-determination procedures will correspond to which sentences in which contexts, we may want to say that the set of sentence-meaning ordered pairs corresponding to a natural language is an *open* one.

Open compositionality does not only tell us what natural languages *are not*, by endorsing the *Lewisian Compromise* the view also gains insight into what natural languages *are* in fact. Chomsky (1986) famously claims that, as students of language, we should focus on its knowledge and ask three important questions: What do we know? How do we come to know it? And how is this knowledge used? By seriously endorsing the *Lewisian Compromise* we can offer an empirically supported answer to each one of these questions. In so

doing, *open compositionality* shares the central goal of *strong compositionality*, namely, to explain how it is that competent speakers manage to determine the meaning of complex expressions.

4.2 LANGUAGE AS A SUPERMODULE: WHAT WE KNOW AND HOW WE KNOW IT

What do we know when we know a language? According to the evidence discussed in chapter 3, when we acquire a natural language we obtain knowledge from multiple domains, such as knowledge of phonemic inventories; knowledge of a relevant vocabulary; knowledge of patterns and regularities; knowledge of grammatical constructions; and knowledge about the statistical distribution of word categories in discourse. We also acquire higher order forms of knowledge, such as knowledge of mental states; knowledge of others' perspective, intentions, and their relevance for communication; and knowledge about the nature of social interaction. In sum, what we know when we know a natural language is knowledge from multiple and varied domains and orders of cognition, including language specific cognition together with domain general knowledge. Knowledge of language is not of a highly specialized kind, but rather a heterogeneous collection of different cognitive domains. Language cognition appears to be the arena where otherwise unrelated areas of cognitive expertise come to interact with each other—e.g., knowledge about the distribution of phonemic patterns, knowledge about the mental states of others, and knowledge about how to identify intentions and consider the perspective of others, may help a speaker identify a novel word, interpret it as a proper name and recognize its intended referent.

Given that there is no language specific cognitive mechanism or domain of knowledge, linguistic cognition characterizes itself for being cognitively penetrable. Non-linguistic kinds of knowledge—e.g., intention understanding—interact with, and even give shape to, linguistic knowledge—i.e., lexical, syntactic, and semantic. This interaction happens all across linguistic cognition, including the linguistic subdomains, with lexical knowledge helping syntax and vice versa. Far from being some kind of modular knowledge holding specialized information about the linguistic domain, knowledge of language appears to be more like a supermodular kind of cognition, where knowledge from distinct, modular and domain general sources of cognition are recruited for a common cognitive task. Knowledge of language is, thus, knowledge of a complex (multiple orders), highly interactive (multidirectional), and penetrable (cross and sub-domain) representational system.

In chapter 3, I argued that, with its endorsement of *strong compositionality*, the *closed* view was also committed to empirically testable cognitive hypothesis, which I called *cognition* and *autonomy* (see section 3.1). Unlike the *closed* view, *open compositionality* does not endorse *strong compositionality*. Thus, it is not committed to the truth of either *cognition* or *autonomy*. Instead, cognitively speaking *open compositionality* is committed to *open* and *interactive* siblings of each of these empirically testable hypotheses. I dub these *open* hypotheses *supermodular* and *interactive*, both of which are a consequence of embracing the evidence presented in chapter 3 (see sections 3.3 to 3.7).

Supermodular: for any speaker *S* and any natural language £, *S* is a competent speaker of £ if and only if *S* possesses and/or exercises a set of multiple and varied, modular and domain general, language specific and non-linguistic, abilities thanks to which *S* acquires, develops and sustains her knowledge of £.

Interactive: for any speaker *S*, any language £ with a lexicon λ, a syntax σ, and a semantics ϕ, *S* is a competent speaker of £ if and only if *S*'s knowledge of λ, σ, and ϕ interact amongst each other, as well as with information from non-linguistic domains such as intention understanding, perspective taking, theory of mind, and statistical analysis among others.

On this view, competent speakers have a wealth of resources available for use, from statistical and distributional analysis to Theory of Mind, mathematical reasoning, moral reasoning, emotion understanding, practical reasoning, general knowledge, and many other modular or domain general cognitive resources. Knowing a natural language is, thus, knowing how to use this supermodular cognitive capacity, equipped with lexical, syntactic, and semantic information, and capable of recruiting and being recruited by multiple other, modular and non-modular, cognitive abilities. How do we come to know, or acquire, such an outstanding form of cognition? As it emerges from the evidence on language acquisition (see Ambridge and Lieven 2011; see also sections 3.3 to 3.6) human subjects appear to gain knowledge of language by means of multiple mechanisms and strategies employing information from multiple sources. There is evidence that subjects make use of fast word-to-world mapping, categorical perception of speech, statistical analysis, pattern recognition, Theory of Mind, learning biases, prelinguistic conceptual biases, social interaction, intention recognition, practical reasoning, and general knowledge, among others. Each one of these is a cognitive mechanism or strategy that, if taken in isolation from the rest, may appear to have little or nothing to do with knowledge of language. It is when they are all put to

work in conjunction to solve a common task that they become the backbone of language acquisition.

Knowledge of language is, as I argued above, knowledge of a highly interactive, multi-domain, representational system allowing us not only to communicate with others, but also to strengthen and develop our cognitive abilities. As a matter of fact, knowledge of language is essential for a child to develop, cognitively speaking, into an adult. This is so in part because the exercise of linguistic competence is the exercise of such a varied and complex set of mechanisms, abilities, strategies, and domains of cognition, and in part because linguistic competence is itself the *product* of such a complex cognitive endeavor. As Shatz (2008) puts it, natural language is both a "consequence and an enabler of the exercise of higher-order" cognition.

4.3 SEMANTICS AS DECISION-MAKING: HOW WE USE IT

With the endorsement of *strong compositionality* the closed view obtained an account of how the meaning of complex expressions is determined. Furthermore, the account is extraordinarily simple and methodologically efficacious and, hence, it can be easily extended to account for the meaning of *any* complex expression in any given natural language. Contrary to this, after a careful study of the available empirical evidence, and in virtue of its endorsement of the *Lewisian Compromise*, the *open view* abandons the hypothesis that natural languages may be *strongly compositional* systems of representation. This naturally prompts the question: if not *strong compositionality*, what then? A different hypothesis about the meaning of complex expressions must take its place, one that is compatible with actual human cognitive and psychological limitations. Thus, our question becomes Chomsky's third: how do we use our supermodular, interactive knowledge of language?

Fortunately, the immense psycholinguistic literature on language processing already includes an answer—or, rather, multiple ones. But before pondering each one of these answers, it is worth asking what theoretical role is there left for semantics within the *open* view of natural languages. According to the orthodoxy (see Frege 1892; Russell 1905; Tarski 1936; Carnap 1947; Montague 1974; Lewis 1970; Stalnaker 1978; Kripke 1980; and Kaplan 1989), semantics, as a theory of meaning, is in the business of determining truth-conditions. As I argued in previous chapters (see sections 1.1, 2.1, and 2.2), this conception of semantics naturally follows from the closed view. Given that the goal is to offer a mechanical, algebraic account of the meaning of any possible sentence of a given language, meanings must be entities of the sort that can be part of an algebraic account (e.g., variables, arguments, or values

of a mathematical function). As Lewis (1970) famously puts it, "(if) we want to say what a meaning *is*, we may first ask what a meaning *does*, and then find something that does that." (Lewis 1970, 193) If we are asking about a sentence, then meaning is that which determines "the conditions under which (it) is true or false." (Ibidem) Or, using a more recent language (owed to Lewis 1975; Stalnaker 1978; and Kaplan 1989) meanings are functions from possible worlds to truth-values. If we are asking about an expression part of a sentence, then meaning is whatever contribution the expression makes to the truth-conditions of the whole sentence. Briefly put, according to the *closed* view of language, semantics is mainly a matter of truth-conditional *calculi*.

It should be clear, then, that the motivation for a truth-conditional understanding of semantics is essentially tied to the endorsement of *strong compositionality* for natural languages. Once we abandon the latter commitment, and with it the very possibility of offering an algebraic account of meaning for any natural language, there is no important reason for truth-conditional *calculi* to play an essential role in the determination of meaning for complex expressions. Lewis (1970) is right in suggesting that to determine what a meaning is we will do well in asking first what a meaning does. When it comes to natural languages we know, *contra* Lewis, that meaning does not compositionally determine truth-conditions, for natural languages do not observe *strong compositionality*. We must, therefore, ask again what a meaning does.

In this new, psychologically informed framework to answer this question we must first ask what competent subjects do when understanding a sentence. The meaning of a sentence is something that competent subjects use as an adequate representation of what the speaker is trying to communicate. Thus, if we want to know what meaning does, we need to know how competent subjects go on interpreting sentences to properly represent others' use of language.

Language use takes place in variously constrained situations from the point of view of *cognition*. Competent subjects usually do not have unlimited time to come up with an interpretation of a given sentence. And aside from time limitations producing cognitive burden, there are cognitive restraints having to do with limited working memory and attention (among others). In this sense, language use and language processing always take place in a context of bounded *human* rationality. Neither speakers nor hearers have vast amounts of time and mental resources to allocate for language use. Both speakers and hearers want to get others to do what they want them to as fast as possible. These circumstances heavily determine what competent subjects can and will do when it comes to using and understanding natural language sentences.

Language use also takes place within contexts substantially constrained by *practical interests*. Competent subjects are always expected to offer an

appropriate response to what has been said, but the appropriateness of the response wildly varies from context to context. Sometimes interpreters are expected to offer a detailed response, other times they are merely expected to engage in some or other motor action, and sometimes even a mere nod will be enough. Yet, the idea of an appropriate response that is contextually defined plays an important role in determining the kind of interpretation that hearers will be likely to produce.

Thus, we get to a first answer about what it is that meaning does. Meaning delivers an appropriate interpretation of a sentence in a context in a way that fits the practical limitations in accordance with the cognitive constraints. With such practical and cognitive limitations it is a mistake to demand a complete and detailed representation for each and every sentence in any context. Such an interpretation would not be an appropriate response in many cases and, most importantly, expecting it would be more than is psychologically reasonable to anticipate from competent subjects. Most of the time a partial, perhaps somewhat incomplete, interpretation will be appropriate. Speakers and hearers should be expected "to do the least amount of work necessary to arrive at a meaning for a sentence." (Ferreira and Patson 2007, 81)

In a series of multiple studies (see Bailey and Ferreira 2003; Christianson, Hollingworth, Halliwell, and Ferreira 2001; Christianson, Williams, Zacks, and Ferreira 2006; Ferreira 2003; see also Ferreira, Ferraro, and Baily 2002; and Ferreira and Patson 2007), Ferreira and her colleagues have carefully studied how adult speakers process sentences of different sorts, including garden-path as well as ordinary sentences. Their results consistently show that language processing is often shallow, yielding incomplete representations as interpretation. Christianson et. al. (2001) show that adults regularly misunderstand the syntactically determined meaning of garden-path sentences, even when given enough time to process. Ferreira (2003) shows that even ordinary sentences may be misunderstood when presented in noncanonical forms. The evidence shows that

> Although the comprehension system clearly makes use of syntactic algorithms, it also uses heuristics, which generally operate more quickly. The heuristic interpretation might then be selected because it becomes available more quickly, or because the system has some reason for preferring it (e.g., it conforms to real world knowledge). (Ferreira and Patson 2007, 74)

A series of studies following event related potentials (ERPs) (see van Herten, Kolk, and Chwilla 2005; Otten and van Berkum 2007; Nieuwland and van Berkum 2005; see Kuperberg 2007; van Berkum, Koornneef, Otten, and Nieuwland 2007; van Berkum, Brown, and Hagoort 1999; see also van Berkum 2008, for a review of ERP studies of language processing) show that

competent speakers commonly anticipate discourse and fix on an interpretation even before they hear the end of the relevant word or sentence. Furthermore, speakers are open to revise the syntax of a sentence in order to obtain a better interpretation. This shows that speakers use something else, something more, than just syntactic algorithms to process natural language. They may use an alternative strategy to arrive at an adequate interpretation more quickly than by following a compositional procedure.

The empirical evidence strongly suggests that the above-mentioned view of meaning is correct. The meaning of a sentence is any interpretation that meets the following two requirements. First, it must be appropriate—according to the contextually variable *practical interests* relevant in the context—and, second, it must be an interpretation that requires little amounts of time and cognitive resources to arrive at. Briefly put, rather than determining truth-conditions, meaning determines an economical (time and cognitive wise) and appropriate (according to contextual standards) interpretation of a sentence.

Now that we know what meaning does, we may look for something that gets this job done. If not by means of truth-conditional *calculi* based on lexical, syntactic, and semantic information, how do competent speakers manage to complete it? Information about the words and syntax of the sentence is important to solve the task; there is no doubt about it. But so are the limitations of appropriateness, available time, and mental resources that interpreters must consider when getting the job done. Ferreira and her colleagues (see Ferreira and Patson 2007) have a fruitful alternative suggestion, to view sentence interpretation a decision-making task.

Ferreira and colleagues underscore one more limitation shaping this semantic decision-making task. Aside from *practical* and *cognitive* limitations there are important *informational* constraints. Subjects typically interpret sentences based on incomplete information about the semantic features of the corresponding sentence. This informational incompleteness is owed to two independent factors. First, speakers cannot obtain full information because getting it, including details about the syntax and semantics of the sentence, requires substantial amounts of time and mental resources that are rarely available. Interpreters would have to wait until the speaker concludes and allocate expensive cognitive resources (e.g., attention, working-memory, etc.) to build a full representation from memory. Interpreters cannot dedicate so much time and resources for every single sentence they encounter. Second, speakers cannot obtain full information because there is typically no such thing as a well defined set of information that fully determines what a sentence means in a context. Even if the speaker can somehow identify all relevant syntactic and semantic details these are typically not sufficient to determine meaning. After all, natural languages are not, on this view, logically closed *strongly compositional* systems of representation.

As a decision-making task, sentence interpretation must be performed under *practical*, *cognitive*, and *informational* limitations. Thus, interpreters are facing the sentence-interpretation task from uncertainty (see Gigerenzer 2008). It is for solving tasks such as these that interpreters may benefit from what Gigerenzer and the ABC Research Group call "fast and frugal heuristics" (see Ferreira 2003). Heuristics are decision-making strategies useful to *find* out or come up with the best decision for a specific problem in a context. These strategies are said to be "fast and frugal" as they do not involve much computation (i.e., fast) and only take into account some of the available information (i.e., frugal) (see Gigerenzer, Todd, and the ABC Research Group 1999). Fast and frugal heuristics fit well with the multiple constraints of sentence interpretation. They require little amounts of time, mental resources, and information. Heuristic decision-making or, in this case, heuristic semantics, does well by ignoring detailed information about the sentences (i.e., ignore part of the sentence's structure) and fixing on an interpretation from early on by roughly picking from a couple options.

The idea of fast and frugal heuristics in semantics, or fast and frugal strategies for finding out an interpretation of a sentence, is well suited to explain how speakers find an appropriate and economical interpretation based on incomplete information. But heuristic strategies are not the only available strategies. In at least some cases the task requires a detailed interpretation of the sentence. In such cases, following a syntactically driven procedure to build an interpretation in a compositional manner may offer the best solution. So how is it that heuristic strategies and syntactically driven processing coordinate to offer a solution for the interpretation task? There are good reasons to think that syntax-driven strategies are exceptionally employed. The most telling one is the fact that they are rather costly, as they require that interpreters obtain full and detailed information about the sentence to be interpreted, hence requiring more time and mental resources. The fact that in many cases the most appropriate and economical interpretation is not the one determined by the relevant syntactic algorithm (see van Berkum, Brown, and Hagoort 1999; Otten and van Berkum 2007), and the fact that interpreters arrive at such interpretation from early on, suggests that syntax-driven strategies are not run in first place with heuristics appearing only if the former fails. The evidence suggests that heuristics may appear from early on in sentence processing.

Thus, there seem to be two alternative ways in which syntactic procedures and grammatical principles may interact with heuristic strategies in order to coordinate a proper interpretation. Either heuristic reasoning runs first, with syntactic algorithms being used only when it fails, or both are run in parallel (see Jackendoff 2007). Ferreira (2003) argues, for example, that subjects use simple heuristics to process sentences first and then "coordinates the output of those heuristics with the products of more rigorous syntactic algorithms."

(Ferreira 2003, 192) (but see Townsend and Bever 2001 for an alternative model). But there are also reasons to think that syntactic / grammatical processing runs in parallel with heuristic reasoning, as evidenced by cases of ellipsis—e.g., John cut the pie and Mary the pancake—where grammatical principles are required to identify the missing clause in "Mary the pancake" by looking back at the syntactic context and obtaining what appears to be the appropriate interpretation—i.e., that Mary *cut* the pancake. If in fact such syntactically based interpretation gets selected from early on in the process—and this can only be determined through empirical testing—then this suggests that syntactic algorithms are not merely run as a backup strategy, but may coordinate with heuristic strategies by running in parallel with them.

So, if a meaning's job is to determine an appropriate and economical interpretation based on incomplete information, and the question is what gets this job done, the answer will be (fast and frugal) decision-making strategies aided by syntactic algorithms working in parallel *or* as backup. The goal of semantics is, then, to find out which specific strategies speakers follow to solve this task in certain contexts and how they do so. Much of this work has already been done in psycholinguistics, identifying relevant heuristic strategies thanks to multiple experimental designs (see Ferreira and Patson 2007; van Berkum 2008). For example, the evidence suggests that competent English speakers follow a common heuristic strategy when interpreting declarative sentences, namely, to assume that the first NP is the agent of the action and the second NP the recipient of it (Bever 1970). Other studies (see Herten et al. 2005; Tabor, Galantucci, and Richardson 2004) suggest that speakers follow plausibility heuristics, producing interpretations of sentences as describing something plausible even at the expense of modifying the sentence's syntactic structure. Surprisingly or not, a great deal of work in contemporary truth-conditional semantics and pragmatics is also relevant to find out how subjects solve their interpretation task. For example, the assumption that ordinary proper names designate one and the same object across counterfactual scenarios—Kripke's (1980) seminal contribution to contemporary semantics—may coordinate with a heuristic strategy—e.g., the assumption that the first NP of a declarative sentence is the agent and the second the patient of an action—to deliver a quick interpretation of a sentence involving modal claims.

A careful study of the immense work done by philosophers of language and linguists within the closed view is needed to determine which proposals are mere consequences of a staunch defense of *strong compositionality*—e.g., Frege's (1892) sense and denotation distinction—which may have an independent motivation—e.g., Kripke's (1980) rigidity for proper names—and which are somewhere in between—e.g., Kaplan's (1989) character and content distinction. It goes without saying that, whichever they may turn out

to be, the independently motivated accounts of meaning for distinct kinds of expressions will be seen as yet one more decision-making strategy available for the task of language interpretation. Thus, they will no longer be seen as *necessary* semantic features of the relevant kind of expression. Language interpretation is an adaptive task. If the external conditions shift radically enough, a change in strategy may be needed. Thus, the fact that a given strategy of interpretation—if you prefer, meaning—proves to be successful with a certain kind of expression for solving several interpretation tasks does not exclude the possibility that a different strategy—if you prefer, a different meaning—may be needed, in relation to what seems to be the same expression, for a different task. A successful language processing strategy is such only in relation to a given task in a given context. Hence, it is fallible, revisable, and not universally applicable.

To sum up, according to *open compositionality* semantics is conceived of as an account of the decision-making process that sustains language processing. Language understanding is a matter of deciding which one is the most appropriate interpretation given economical, practical, and informational limitations. Something similar may be expected to happen from the point of view of language *production*. Speakers face a decision-making task, namely, that of finding the most appropriate sentence to convey her thoughts given the economical, practical, and informational limitations of the context. Fast and frugal heuristics aided by syntactic algorithms are well suited to help solve this task as well.

Let me conclude this section by describing a few salient features of semantics and meaning according to the *open view* I am proposing. With respect to semantics in general, by endorsing the idea of semantics as decision-making, we lose the goal of an algebraic, or formal, account of meaning for all complex expressions of a given natural language. Semantics—i.e., the theory of meaning for complex expressions in a natural language—becomes a deeply empirical matter. Competent subjects need not have *a priori* access to the set of heuristic strategies at their disposal for solving everyday semantic decision-making tasks. As a matter of fact, subjects usually ignore this aspect of ordinary human rationality. And there is a good explanation for this. Heuristic decision-making is an adaptive form of rationality, essentially relating human thinking with its environment (see Gigerenzer 2008, on ecological rationality), given the goal of finding which strategy best fits the constraints of this or that particular situation. Heuristics are, thus, externally determined strategies, the success / adequacy of which depends on environmental conditions. Therefore, on this view of semantics, to find out exactly how it is that humans competently interpret sentences, we need empirical testing. As a theory of how complex expressions get their appropriate and economical interpretation semantics is an *a posteriori* enterprise.

The account offered has three important further consequences. First, it allows for ranks, ties, and conflicts. There may be different interpretations of a single sentence that are appropriate and economical to different degrees. If such is the case, there may be a ranking of interpretations with the most appropriate and economical interpretation on top. There may be conflict between the appropriateness and the economy of the interpretation. There may also be two or more interpretations of a single sentence that may be equally appropriate and economical in the context. This is as it should be. Once we abandon the project of an algebraic account we should expect more than one meaning, or interpretation, corresponding to a sentence.

Second, the meaning of a sentence may vary from context to context. Even if the sentence has no obvious indexical or context dependent expression, the meaning of the sentence may fluctuate in virtue of variations of the appropriateness condition, perhaps because there is a different amount of time available or there are different practical interests. These variations are well known by philosophers, especially by those on the "pragmatic view" as Travis (1997) dubs it. Following Travis's own example, the sentence "The leaves are green" means different things depending on whether we are in a context where participants are interested in issues of interior design, or in issues of botany. The interpretation of a sentence may also vary depending on how economical it is given the resource restrictions of the situation. A sentence's meaning—its appropriate interpretation—may vary depending on how much time and cognitive resources we are given as interpreters.

Third, and last, the meaning of a sentence may in some cases be determined in a syntactically driven manner. There is no principled reason to exclude the possibility that a syntactically guided, compositional combination of the meaning of the expression parts of a sentence delivers an interpretation that happens to be the most appropriate and economical one. In other words, on this view something like compositional processing is an available strategy. It is just no longer the only one available.

These three features of meaning according to the *open* view—i.e., allowing for ranks, ties, and conflicts; allowing for different meanings under different practical, economical, and informational conditions; and allowing for compositional processing in different ways—illustrate the *openness* and *compositionality* of *open compositionality*.

4.4 THE COGNITION-FIRST METHODOLOGY

Open compositionality offers a psychologically based and cognitively enhanced view of natural language. This view is not only more accurate, as it is fully supported by the empirical evidence (see chapter 3 and sections 4.2

and 4.3 of this chapter), it is also more credible than the account offered by the closed view based on *strong compositionality*. Whereas open compositionality offers an account of language that is tailor-made for actual human cognitive resources, the closed view offers an account that is suitable for logical-mathematical analysis. As should be expected, there is a price to be paid for having such an empirically adequate account. With open compositionality we lose almost all the benefits of the closed view. I have already mentioned one of them, the algebraic account of meaning. The project of determining the meaning of any possible sentence of a natural language becomes less tractable—if not implausible—as it transforms into a deeply empirical, externally constrained, and context dependent issue. The closed view also comes with two more outstanding features, namely, its simple and efficient methodology (see sections 1.1, 2.2, and 2.3) and its ingenuous account of the productivity and systematicity of natural language (see section 2.4). These outstanding components are gone once we move towards open compositionality. The new methodology is substantially more complex and unspecific; and its account of productivity and systematicity is more nuanced and constrained. I will deal with these issues in this and the following section. As I will argue, the resulting methodology is worth keeping, and the accounts of productivity and systematicity are better suited to the *explanandum*.

Aside from its alleged explanatory power, the methodology of *strong compositionality* is outstanding for its simplicity, transparency, and its applicability. It is surprisingly *simple* as there are three variables to control and /or take into account, i.e., lexicon, syntax, and semantics. It is remarkably *transparent* in that it delivers precise instructions to the theorist on how to proceed for each particular case, given that there are at most three possible theoretical moves. And, consequently, this simple and surprisingly precise methodology can be directly applied to any possible problematic case, no matter the context, intensions or practical limitations, and no matter what kind of simple or complex expression we might be dealing with. The resulting theory may, of course, be quite complex (see Elbourne 2005, on definite descriptions) but the methodology stays simple, precise, and directly applicable.

These exceptional features are, of course, dependent on the assumption that natural languages observe *strong compositionality*. The methodological simplicity depends on there not being any non-linguistic information relevant for the determination of meaning for complex expressions. Language use and processing must be cognitively impenetrable by other domains of cognition. Any interference of information from other mechanisms or domains will immediately open (i.e., beyond lexicon, syntax, and semantics) the set of relevant variables to consider. Something similar happens to the transparency and precision of the methodology. If natural languages are open to interaction with non-linguistic domains of cognition, the set of alternative theoretical

moves one may follow opens up, and depending on what sort of information or cognitive mechanism intervenes the relevant options may differ in kind from each other. Thus, there is no predetermined set of explanatory moves and it would become impossible to deliver precise instructions on what to do to account for any possible problematic case. As theorists, we would have to find this out in a case-by-case basis. Consequently there could be no precise and simple methodology applicable to all possible problematic cases.

Given that natural languages do not observe *strong compositionality* we are forced to abandon this simple, transparent, and easily applicable methodology in favor of a complex and opaque methodology that is only applicable to all cases in a very general fashion, yet one that is compatible with the cognitive complexity of linguistic cognition. Thus, it is true that we lose an extraordinarily attractive methodology by endorsing *open compositionality*. But that is acceptable since we have to let go that methodology anyway, regardless of whether we endorse *open compositionality* or some other view.

So what sort of methodology should we follow, especially as philosophers of language, to better understand an *open* representational system useful for both communication *and* cognition (see section 1.1)? To properly answer this question it is important to take into account what we now know about this open representational system. We know, first, that knowledge of language appears to be supermodular and highly interactive. Knowledge of natural language appears to consist of a special way in which distinct pieces of knowledge (e.g., practical and general), domains of cognition (e.g., linguistic, social, intentional, etc.), and mechanisms (e.g., Theory of Mind, perception, etc.) are coordinated to work for a single task. We also know, second, that competent subjects use this rather distinct cognitive system by means of a mixture of fast and frugal decision-making and syntactically driven interpretations.

It follows from this account of what we know when we know a language and how we use this knowledge, that there is no such thing as a purely linguistic phenomenon to be studied. Natural language turns out to be something more than *just* pure language. So an important methodological principle when studying problematic aspects of natural language would be to avoid focusing only on purely linguistic features such as syntax, semantics, and lexicon. For what lies underneath such problematic aspects of language use is nothing but an interweaving of multiple non-linguistic mechanisms and domains of cognition. Thus, we have a first methodological lesson. If we want to understand natural language we must first understand the cognitive machinery it involves.

Now, a second methodological lesson comes from considering the very goal of philosophical theorizing about natural language, namely, understanding linguistic practice. The closed view aims at explaining linguistic

communication by means of two central assumptions, *strong compositionality* and *correspondence*. According to the latter there is a correspondence between the content of a speaker's thought(s) and the meaning of the sentence(s) she uses. This correspondence is partly explained by the conventions that give place to a particular linguistic practice among the members of a population (see Lewis 1975), and partly by the *strong compositionality* of natural language. On the closed view, there is a one-to-one correspondence between mental content and linguistic meaning, in virtue of them having equivalent compositionally determined truth-conditional content. So linguistic practice is possible because competent subjects have an easy way to express and access to each other's thoughts; they simply produce / process sentences observing *strong compositionality*. Language use, on the traditional view, offers a perfect insight into the mental life of speakers. We can find out exactly what human subjects think by simply looking at what they say.

Open compositionality aims at improving the closed view with a psychologically based and cognitively enhanced theory. The resulting view, however, can only keep the *correspondence* assumption of the *communicative view* without *strong compositionality*. As in the closed view, such correspondence is partly explained by the relevant conventions observed in linguistic populations. After considering the empirical evidence on language processing, we now know that a peculiar mixture of heuristics and syntactic interpretation completes the explanation. It is thanks to successful decision-making strategies that a speaker manages to pick the sentence that, according to the relevant conventions in her community, best corresponds with her thoughts.

The account of linguistic practice remains somewhat the same as before. Speakers get others to understand what they want them to understand thanks to the correspondence between the meaning of sentences and the content of thoughts. There is, however, a wrinkle. Whereas before we had a strong, one-to-one correspondence between mental content and linguistic meaning, in virtue of them having equivalent, compositionally determined content, we now have a *weak correspondence* mediated by a variable decision-making process, both on the side of the speaker producing the sentence and of the hearer interpreting it. Both may, and perhaps most probably will, be the result of heuristic decision-making, quickly selecting the most appropriate and economical sentence / meaning for the purpose at hand. Thus, there is no guarantee that the sentence produced by the speaker offers a complete and detailed representation of the speaker's thought. And there is no guarantee that the interpretation produced by the hearer offers a complete and detailed representation of the sentence chosen by the speaker to express her thought.

The correspondence between the speaker's mental content and the hearer's interpretation is, at best, mediated by two degrees of appropriateness. The

interpreter's representation of the sentence is an appropriate representation of the speaker's sentence, which is itself an appropriate representation of the speaker's thought(s). This weak correspondence does not preclude competent subjects from successfully communicating, but it does preclude us theorists from deriving a detailed account of the content and structure of our thoughts by merely looking into language use. Language use, on this *open* view, does not offer a complete and transparent view into the mental life of speakers. Linguistic expressions merely offer an indirect reflection. So we reach our second methodological lesson: if the goal is to get as close as possible to human thinking we will do better in looking for more resources than just sentence meaning.

Together, these methodological lessons offer a substantial piece of theoretical guidance. According to the first lesson, if we want to understand certain aspects of natural language we must understand the underlying cognitive machinery. According to the second lesson, if we want to understand human thought we must avoid (only) looking into language. Thus, if *qua* philosophers we want to understand natural language, problematic aspects of its use in particular, we should initially ignore language itself and begin by looking directly into human cognition—with the help of cognitive psychology and other cognitive sciences, of course. This delivers what I call the *cognition-first* methodology:

Cognition-first: To understand a given linguistic phenomenon and how it takes place, we must first understand the underlying cognitive processes, as understood by our best cognitive psychological and psycholinguistic theories (among other cognitive scientific endeavors). If we want to understand how X-discourse works, we must start by understanding X-like cognition.

What does cognition-first entail? Assuming that underlying cognitive processes accompany all linguistic phenomena, to put *cognition-first* in the study of a certain aspect of language use is to prioritize the understanding of such cognitive machinery before focusing on language use itself. Such cognitive scientific understanding will offer the necessary elements with which (and from which) any account of language use should work. Cognition, according to this methodology, constitutes the cornerstone of our study of language. It sets the limits, the *desiderata*, for any satisfactory account. In following the *cognition-first* methodology the open view inverts the explanatory order of the closed view. According to the latter ordinary language use by competent speakers is a transparent reflection of the speaker's thoughts. This transparency is specially well justified given that both, thought and language use, are said to have contents that correspond to each other in a strong sense. On the open view it is thought the one that offers an access into the nature of language

use, and this access is no longer guaranteed by an assumed correspondence of contents but by independent empirical evidence. This methodology offers a language-independent access to the nature of human thought and cognition. In virtue of this we may be capable of achieving a better understanding of natural language if only because language is an expression of thought and not the other way around. Aside from these empirical advantages, the new methodology guarantees that our theories observe the *Lewisian Compromise* by describing a language that can, in a strict sense, be used by human beings.

To illustrate, consider a specific theoretical goal. If we want to understand *reference* in natural languages, we may start by looking at studies on reference resolution (see Baldwin 1991; Hanna, Tannenhaus, and Trueswell 2003); and if we want to figure out how proper names work, we may want to look at studies on the acquisition of proper names, their corresponding cognitive architecture, and the nature of memory dedicated to proper names (see García-Ramírez and Shatz 2011). Failure to follow this methodology may result in cognitively implausible linguistic theories. Possible-worlds accounts of modals and counterfactuals (see Stalnaker 1968; Lewis 1973; and Kratzer 2012) for example, claim that to understand such expressions, competent speakers must consider alternative possibilities given by logically possible worlds. Yet it is unclear that subjects consider anything like logically possible worlds when engaging in modal reasoning. Empirical studies (see Byrne 2016) show that understanding counterfactuals requires the use of imagination, not so much logical reasoning, by considering iconic (not truth-conditional) mental models as alternative possibilities (see also Byrne 2005).

The *cognition-first* methodology, however, does have certain disadvantages. Unlike the *compositionally closed* methodology, it is far from being simple and it does not come equipped with precise instructions on how to proceed to account for a given problematic phenomenon. Consequently, it is applicable only in a general way. *Cognition-first* tells us to put an understanding of cognition first in the explanatory order. But it does not tell us where to look for such an understanding or how to achieve it. The best we can do as students of language is to follow the lead of cognitive scientists, particularly cognitive psychologists and psycholinguists. Doing so will not be easy, as it presupposes more work and study to achieve familiarity and knowledge of the disciplines. The price, however, is worth paying, as I will show in the remaining chapters of this book where I will offer novel accounts of distinct problematic phenomena by following the *proposed* methodology. Chapter 5 offers an account of what philosophers know as the problem of "substitution failure," chapter 6 deals with the so-called problem of empty names, and chapter 7 advances a novel understanding of moral discourse in terms of moral knowledge and cognition.

4.5 PRODUCTIVITY, SYSTEMATICITY, AND COMPUTABILITY

Productivity, systematicity, and computability are commonly viewed as evidence of *strong compositionality*, as only by assuming the latter can we account for the former, or so it is argued. However, it has been observed (see section 2.4; see also Szabó 2012; Baggio, Lambalgen, and Hagoort 2012) that neither productivity nor systematicity or computability offer direct support for *strong compositionality*, as they may be accounted for by assuming much weaker hypotheses. What accounts for *productivity, systematicity,* and *computability* according to *open compositionality*? As I will show in this section, the understanding of semantics as decision-making offered in section 4.3, with its account of language processing as a mix of heuristics and syntactic algorithms, does offer a satisfactory account.

Natural languages are *productive*. They have an enormous (perhaps infinite) amount of sentences and an equally great amount of meanings corresponding to each sentence. Human subjects are cognitively limited. Their memory has a finite storage capacity, clearly not big enough to store an enormous (or, worse, infinite) amount of sentences and their corresponding meanings. Yet, human subjects manage to acquire and competently develop natural languages, and so they are capable of knowing all possible sentences and understand all possible corresponding meanings. The challenge is, thus, to explain how limited beings can be *capable* of knowing an unlimited amount of sentences and their meanings without *actually* knowing all of them.

In section 2.4, I argued that *productivity* is far from offering support for *strong compositionality*, as it only demands that there be a finite procedure by means of which competent subjects interpret any possible sentence, for an unlimited (perhaps infinite) number of them. Of course, *strong compositionality* satisfies the demand by means of an algebraic, compositionally closed, procedure based on limited lexical, syntactic, and semantic knowledge. Yet, *productivity* is far from demanding an algebraic and compositionally strong procedure. Any finite procedure will do, so long as it works for every possible sentence. As Szabó (2012) argues, *productivity* is, at best, evidence of *weak compositionality* (see sections 2.2 and 2.4). In fact, it can be accounted for by an even *weaker* principle (see section 2.4).

Weaker Compositionality: The meaning of a compound expression is *partly* a function of the meanings of its parts and of the way they are syntactically combined; and *partly* a function of *domain general* influences and other relevant human cognitive limitations.

After a detailed review of the empirical evidence on language acquisition (both lexicon and syntax) and language processing (see chapter 3, sections 3.3 to 3.7) a more accurate, independently supported account of *productivity* emerges. On the one hand, the evidence shows that knowledge of language requires an interweaving of multiple mechanisms, domains, and sources of information, including both linguistic and non-linguistic information—i.e., intention understanding, statistical analysis, general knowledge, practical knowledge, ToMM, etc. On the other hand, the evidence also shows that language processing is a highly resource-constrained task that is typically solved with the help of heuristic strategies and syntactically driven algorithms. Together, the supermodular view of knowledge of language and the decision-making view of language processing deliver a satisfactory account of natural language *productivity*.

First, both make very limited cognitive demands from subjects. The supermodular view of linguistic knowledge does not add any special memory requirements, as it only constitutes an interweaving of knowledge or cognition that is already granted for multiple other purposes. And the view of language processing as decision-making is even less computationally demanding than *strong compositionality* as it relies on little computation (fast) and only partial information (frugal). Second, both the supermodular knowledge of language and the decision-making strategies involved in language processing are general in scope, as none of them involves knowledge or processing instructions for specific sentences in specific contexts. The result is a finite knowledge capable of running a finite decision-making procedure to find an appropriate and economical interpretation for any given sentence in any given context. This is what accounts for the *productivity* of natural language. It seems appropriate to call this the principle of *open compositionality*.

Open Compositionality: The meaning of a compound expression *E* of £ is always the result of a decision-making procedure for finding out the most appropriate and economical interpretation of *E*. Depending on contextual demands, this procedure may sometimes involve heuristic strategies, syntactic algorithms, or both as they are run in parallel or one as a backup strategy of the other.

If natural languages observe *open compositionality*, competent speakers need not learn and store an unlimited amount of different sentences associated with their meanings in order to find an appropriate and economical interpretation of it. All that speakers need to learn / develop is lexical, syntactic, and semantic knowledge interweaved with a host other multiple mechanisms, domains, and sources of information—some of which must also be learned / developed. With enough expertise, speakers will acquire decision-making

strategies allowing them to find appropriate and economical interpretations for any given sentence.

Let me now consider *systematicity*. Natural languages are said to be *systematic* because there appear to be relations between the meaning of some complex expressions and that of other complex expressions sharing the same syntactic combination. Thus, any competent speaker who understands, say, expressions E_1 and E_2, made up of distinct expression parts but identical syntactic structure, will be able to understand any other series of complex expressions $E_3, \ldots E_n$, made up of yet another set of distinct expression parts as those of E_1 and E_2 but identical syntactic structure. *Systematicity* poses a different challenge to the one posed by *productivity*. There is no tension between human limited cognition and the unlimited expressivity of language. Instead *systematicity* shows that the meaning of expression parts and the syntax of a given complex expression *do* play an important role in determining its meaning.

Strong compositionality, of course, offers a direct account of *systematicity*. According to the former, lexicon and syntax are necessary elements in the determination of meaning for all possible complex expressions of a given language, as they are part of a logically closed semantic algebra. But there is no need to endorse *strong compositionality* in order to give lexical and syntactic knowledge the place they deserve. It will be enough to have them both, lexical and syntactic knowledge (and also semantic knowledge), without the idea of them being part of a strongly compositional system.

Open compositionality offers an account such as required. On this view, speakers always draw from their lexical, syntactic, and semantic knowledge, as well as from a host of multiple other sources, to produce and understand sentences. It is not a surprise, for the *open view*, that whoever understands a set of complex expressions sharing a common syntactic structure *can* understand further complex expressions with the same syntactic structure. If this is what *systematicity* demands—i.e., an explanation of how competent speakers manage to understand sentences by means of lexical and syntactic information—then the challenge is already met, as *open compositionality* explicitly considers the use of syntactic algorithms as alternative interpretation procedures. *Open compositionality* does not exclude syntactically driven interpretations; it only adds the use of heuristic strategies as an alternative. Thus, merely understanding *E1* and *E2* allows competent speakers to understand other expressions *E3, . . . En* with the same syntax because, according to *open compositionality*, competent speakers can follow syntactic algorithms to interpret complex expressions. Natural languages need not observe strong compositionality in order for them to be *systematic*.

Finally, let me consider the third and last piece of alleged evidence of *strong compositionality*. Aside from being productive and systematic, natural

languages are evidently *computable*. As Pagin (2012) puts it, "we manage to convey new contents by means of new sentences *in real time*, that is, that a hearer manages to compute the meaning online of an uttered sentence at a speed that matches the speed of speech." (Pagin 2012, 510) Pagin claims that real time interpretation of language requires that the latter be *computable* in an importantly restricted sense according to which it involves relatively few and simple steps. *Strong compositionality*, it is argued, describes a series of few and simple steps required to interpret any possible sentence. According to *strong compositionality* all that speakers need to do in order to interpret a sentence is to identify the lexical meanings and the syntactic structure in order to apply the relevant semantic combination rules.

It is disputable whether the *computability* of language requires something like *strong compositionality*. First, it is clear that the correct sense of *computability* must be given by empirical research on human language processing. If natural language *computability* is to be an interesting notion it must be understood in terms of human cognitive resources. Natural languages are computable *for humans.* Second, as Szabó (2012) claims, what real time interpretation of sentences shows is that whichever information interpreters may need it is readily available as soon as the sentence is uttered, it does not show that such information is limited to lexical, syntactic, and semantic information. That said, I think there is a more conclusive reason to think that real time computation does not support *strong compositionality*. In fact, as I will show, it goes against it.

Let us grant Pagin (2012) both that *strong compositionality* is humanly computable and that whenever hearers have an interpretation available as soon as a sentence has been completely uttered it is because only lexical, syntactic, and semantic information is being computed. The problem for Pagin's argument is that real time interpretation of sentences is much faster than what *strong compositionality* can offer. The evidence on language processing (see van Berkum 2008) shows that competent subjects fix on an interpretation of a sentence *even before* the whole sentence has been uttered. This demonstrably shows that competent subjects are following a computation that is simpler and involves fewer steps than those required by *strong compositionality*. It also shows that, at least sometimes, competent subjects do not care to process the syntax of a sentence in order to interpret it, contrary to what *strong compositionality* demands.

Far from confirming it, real time interpretation refutes *strong compositionality*. Now, *open compositionality* does have an explanation, and a very simple one, of real time sentence interpretation. According to this view, sentences may be interpreted by means of fast and frugal heuristics based on partial information and involving very few computations. Using heuristic strategies to find an appropriate and economical interpretation of a

sentence allows speakers to interpret sentences even before they have been completely uttered. Thus, if anything, real time interpretation supports *open compositionality*.

Contrary to what is commonly accepted among philosophers of language, assuming *strong compositionality* is not necessary to explain the productivity or systematicity of natural languages; a much weaker hypothesis, such as *open compositionality* will do. A recent argument from Pagin (2012) claiming that the computability of natural languages supports *strong compositionality* turns out to be counterproductive, as the empirical evidence on language processing shows that natural languages demand faster and simpler computations than the those offered by *strong compositionality*. With no solid argument for *strong compositionality*, the closed view of natural language can let go this problematic, empirically false hypothesis and endorse *open compositionality*. Doing so will not only deliver a satisfactory explanation, it will deliver an empirically supported account that abides by the *Lewisian Compromise*. Putting it in Lewisian terms, a language that observes *open compositionality* is a language that could possibly, in a very strict sense, be learned, developed, and used by human beings.

4.6 OPEN COMPOSITIONALITY AGAINST ALTERNATIVES

According to open compositionality, meaning determines the most appropriate and economical interpretation of a sentence. Thus, I argued, semantics is better viewed as a decision-making procedure the goal of which is to find the interpretation that best fits the economical, practical, and informational limitations of the situation (see section 4.3). Semantics is, thus, not in the business of determining truth-conditions, at least not as its main goal—thought it may be that doing so delivers an appropriate interpretation in the context. This view appears to be similar to other available proposals that are also based on a rejection of truth-conditional semantics. So before closing this chapter let me briefly consider those proposals in order to show how greatly they differ from open compositionality.

Recanati (2010) and (2012) argues that the traditional *modular* view of semantics, according to which only linguistic information plays a role in the compositional determination of meaning for complex expressions, is flawed. To properly understand how it is that complex expressions get their meanings we need to accept that "pragmatics and semantics *do* mix in fixing truth-conditional content." (Recanati 2010, 3) The traditional "modular" view of semantics is what I have been calling "truth-conditional semantics." The view fails, according to Recanati (2010), and with him many others,

because semantics by itself is not enough to determine truth-conditions, as the latter heavily depend on pragmatic factors. Pragmatic factors are, according to Recanati (2010), contextual elements having to do with the intentions of the speaker. Once we take pragmatic factors into account, we can obtain a full-fledged truth-conditional account of meaning for complex expressions. This account is what Recanati (2010) calls "Truth-Conditional Pragmatics." Truth-Conditional Pragmatics claims that truth-conditional content may be determined by bottom-up as well as top-down pragmatic factors.

It should be clear that open compositionality differs substantially from Truth-Conditional Pragmatics. To begin with, Truth-Conditional Pragmatics is still relying on a *strongly compositional* view of language, one that can deliver an algebraic account of meaning for every possible complex expression. Thus, truth-conditional meaning is still the target. Put differently, Truth-Conditional Pragmatics only differs from the closed view with respect to the complete set of arguments for the algebraic account. The latter view claims it is only lexical, syntactic, and semantic information. The former claims that speaker's intentions must be included too. As Recanati puts it, with Truth-Conditional Pragmatics we can still achieve compositionality: "Take the modulated meaning of the complex: we can say that it is a function of the modulated meanings of the parts (and the way they are put together) *plus, in addition, the context which determines how the content of the whole is itself modulated*." (Recanati 2012, 190)

As such, Truth-Conditional Pragmatics does not account for the highly interactive and *open* nature of human languages (see section 4.2). Following Recanati's description, Truth-Conditional Pragmatics merely helps us move from a modular view of meaning as including only semantics and syntax to a different yet still modular view of meaning as including only semantics, syntax, and pragmatics (i.e., speaker meaning). Furthermore, it fails to account for the evidence on language processing showing that competent speakers interpret sentences in anticipatory ways, without complete information about how the expression parts are combined and, hence, without necessarily following an algorithmic compositional procedure, not even a pragmatically modulated one. From the point of view of open compositionality, Truth-Conditional Pragmatics is just the same as the closed view, albeit with a pragmatic twist.

Sperber and Wilson (1986) and Wilson and Sperber (2002) develop a framework for understanding linguistic communication based on a cognitive interpretation of a generally Gricean inferential model known as "relevance theory." Like the Gricean model (see Grice 1967; 1989), relevance theory claims that linguistic communication is a matter of communicating intentions by inferential means. Unlike the Gricean model, relevance theory has no place for truth among the assumptions needed to achieve such goal. Instead

of truth relevance theory claims that linguistic communication takes place under the assumption of relevance. Relevance here is understood in a technical sense. A piece of information is said to be relevant in two different ways. First, a piece of information is more relevant than another if it does a better job at improving a subject's understanding of the world, if it entails a cognitive benefit, by giving her new information or modifying an old one she may have. Second, a piece of information is more relevant than another if it is less costly processing-wise.

Relevance theory understands linguistic communication as an information-processing task the goal of which is to improve our understanding of the world in the most cognitively efficient way. Linguistic communication is also understood as ostensive behavior, "behavior which makes manifest an intention to make something manifest." (Sperber and Wilson 1995, 49) In other words, linguistic communication is a kind of behavior that aims at improving our understanding of the world by explicitly showing that it is an intentional form of behavior the goal of which is to share intentions. On this view, every act of linguistic communication carries its own intention (or assumption) of relevance. Making this intention manifest to others is essential to language use; its recognition is necessary for the hearer to pay attention to other intentions that the speaker wishes to communicate. This is so because information processing is a cognitively demanding task and hearers will set upon it only if they expect some reward, such as the relevance of the information obtained.

This model of linguistic communication also delivers an account of meaning for complex expressions. On this view, meaning does not determine truth-conditions, but it does determine relevance. Thus, the meaning of a complex expression *E* of a language £ is its most relevant interpretation in the context of communication. This understanding of meaning and linguistic communication is, of course, far from the traditional compositional understanding of language and closer to that of open compositionality. According to Relevance Theory, there is a gap between the compositionally determined interpretation of a sentence, as used by a competent speaker in a context, and its correct (i.e., relevant) interpretation. So Relevance Theory and open compositionality agree that natural languages do not observe *strong compositionality*.

At first glance it would seem that Relevance Theory and open compositionality have further substantial claims in common. The understanding of sentence meaning as relevant interpretation seems not only compatible but also even complementary to the view of meaning as appropriate and economical interpretation under informational limitations here described. Both views understand sentence interpretation as a resource-demanding task whose goal is in part that of easily solving the task. Yet, similarities notwithstanding, there are important and substantial differences between the two views.

First of all, appropriateness is not relevance. There may be contexts where the most appropriate interpretation does not deliver any information that may expand the subject's understanding of the world, or information entailing a cognitive benefit for the subject. In such contexts there is still an appropriate interpretation, even though there is no relevant one. It may be, for example, that the most appropriate interpretation in a given situation is simply a polite one, or it may be that speaker and hearer engage in linguistic communication merely for the sake of it, with no purpose of sharing any relevant information—as when people salute each other. In cases such as these relevance theory and open compositionality will deliver different results.

Second, the processing limitations seem to be substantially different between both views. There appear to be no fixed time or storage limitations for the interpretation procedure to conclude according to relevance theory. So long as the interpretation delivers a grand enough benefit for the subject's understanding of the world, one that outweighs the cognitive cost of processing it, the procedure may continue. This possibility is excluded by open compositionality. Short-term memory is needed for language processing and its storage and temporal limitations are fixed no matter how beneficial for the subject's world-knowledge the interpretation may turn out to be. So, for any such cases where the interpretation procedure must continue well beyond the limits of short-term memory, well beyond the limits of human attention span, open compositionality predicts that the procedure will either deliver a faster, easier and more economical interpretation with lesser cognitive benefit for the subject's knowledge or no interpretation at all. It is not clear, however, what relevance theory has to say about these cases. What does seem clear is that the two views work under different constraints, with open compositionality being explicitly restricted by basic human limitations.

Third, the communication model used by relevance theory takes a complete utterance, say of a sentence, as the basic unit of interpretation. Speakers select a complex expression E from £, made up of expression parts $<e_1, e_2, \ldots e_n>$, because it is the most relevant—i.e, the most beneficial and less effortful—way to *encode* the information the speaker wants to share. Furthermore, they do so because all linguistic communication works on a two-sided assumption of so-called optimal relevance (see Sperber and Wilson 1995, 164; Wilson and Sperber 2002, 604). It is assumed, first, that whatever the speaker wants to communicate by uttering E is relevant enough for E to be worth processing. Second, it is also assumed that the speaker's utterance of E is the most relevant one to communicate the corresponding information, given the speaker's abilities. If this is true then speakers cannot fix on an interpretation of a speaker's utterance before they have complete access to it. If the two-sided assumption of relevance is in place, interpreters should wait at least until the speaker has concluded eliciting her utterance of E before

forming an interpretation that verifies the assumption. This, however, is not always the case. Studies on actual language processing by competent subjects (see sections 3.7 and 4.3) show that many times interpreters anticipate what the speaker will say and fix on a given interpretation before the speaker has concluded eliciting her utterance. Thus, competent interpreters, in many cases at least (see van Berkum 2008), do not seem to care much (consciously or not) about whether the speaker's selection *E* is the most relevant one as described above. In short, for open compositionality sentences may be interpreted with incomplete information about the target sentence, this is not the case for relevance theory.

Finally, perhaps the most telling difference between relevance theory and open compositionality is the former's focus on information processing. Relevance theory correctly admits that natural languages are *not* essentially communicative systems. "Languages are indispensable not for communication, but for information processing; this is their essential function." (Sperber and Wilson 1995, 172) Information processing, however, is far from being a unique function of natural languages. All human cognition, no matter how basic, may be said to have an information processing function. The visual system, for example, is fundamentally an information processing system receiving photons and delivering visual representations.

In stark contrast to relevance theory, *open compositionality* offers a view of natural languages as systems that enable human cognition—or human thinking, if you prefer—because they come equipped with a supermodular interactive architecture. Unlike basic cognitive systems, such as the visual one, natural language enables the sharing and processing of information between multiple cognitive systems, higher order and basic ones. By integrating human cognition in such a dynamic way, natural language becomes an extraordinary platform for human thought to flourish. To quote Shatz (2008) once again, language is both "a consequence and an enabler of the exercise of higher order" cognition. If we view natural language as a supermodular cognitive platform we will be able to understand a great deal of human unique cognition, as well as some linguistic behavior that would otherwise appear enigmatic, paradoxical, or inexplicable. To properly understand certain problematic aspects of linguistic practice we need to recall that language is sometimes playing both roles at the same time, working as a representational system for communicative purposes and as a cognitive platform for higher order cognition.

Finally, the last alternative theory worth considering here is the so-called "allosemy view" of language (see Marantz 2010 and 2013; Wood and Marantz 2017). The allosemy view is part of the more general approach to generative grammar known as Distributed Morphology (see section 3.3). According to distributed morphology there is no place for a traditional lexicon since words

themselves have internal syntactic structure. Complex expressions, including words, are built up from morphological operations. There is, thus, no difference between sentences and words, all the way down to morphologically basic elements such as the root of a word. Within this general framework the allosemy view claims that there is no fixed semantic meanings assigned at the morphological level. Thus, any given root of a word may have multiple semantic realizations or, in other words, it can have multiple different meanings and only the syntactic context determines which is to be selected.

There is no such thing as a single intrinsic meaning, not for words and not for morphemes, on the allosemy view. Given the assumptions of distributed morphology, this entails that there is no single meaning that corresponds to a sentence, as such complex expressions are nothing more than morphological combinations. One and the same syntactic construction can have different meanings. On this view, syntax is completely independent from semantics, while the latter is itself dependent on syntactic locality. As is the case with distributed morphology, the allosemy view is incompatible with *strong compositionality* and, particularly, with the idea that a *sentence* can be given a fixed set of truth-conditions as its meaning independently of the (syntactic) context in which it may appear. Sentences, it may be said on this view, are polysemic in virtue of morphological allosemy.

Even though they both reject the project of a logically closed truth-conditional semantics, as well as *strong compositionality*, the allosemy view and open compositionality differ substantially. To begin with, the allosemy view is meant to offer an advancement of semantics within the more general view of generative linguistics. As the reader may recall (see section 3.3) the latter view does not consider any knowledge beyond syntax, vocabulary (in this case morphemes and their combinations), and semantics to be of relevance. If the allosemy view differs from the closed view it is not because it includes non-compositional elements or processes to be determinants of meaning but, rather, because it requires having much more structure and, hence, much more syntactic composition than considered by the closed view, so much that it ends up not determining any unique semantic interpretation. Yet, unlike open compositionality, the allosemy view does not consider any extra-linguistic information or domain of knowledge—e.g., Theory of Mind, perspective taking, statistical analysis, distributional analysis, and intention understanding, among others—to be of semantic relevance. Even worse, the allosemy view does not even consider the possibility of there being non-syntactic processes—e.g., heuristic strategies—determining the meaning of a sentence. This is a huge and substantial difference, with respect to language processing / meaning determination, between the Allosemy View and open compositionality.

Not surprisingly, there is also a substantial difference, this time with respect to language cognition, between the allosemy view and open compositionality. As any other theory that may be considered part of generative grammar, the allosemy view seems to assume that knowledge of natural language is a highly specialized, domain specific form of cognition. On this view, natural languages are, first and foremost, syntactically driven systems of representation. The allosemy view, in fact, distinguishes itself from other views within generative linguistics precisely because it aims at having a "pure" syntax, completely independent from semantics. Thus, on the allosemy view knowledge of language appears to be modularized, specific, and impenetrable, with syntactic knowledge determining pretty much everything else without being so determined by any other domains. It is hard to describe any account of linguistic cognition that may offer a harsher contrast against open compositionality and its supermodular, highly interactive account (see section 4.2). As such, the allosemy view, at least at first glance, appears to be unable to account for the empirical evidence showing that there is no specific cognitive domain for linguistic cognition (sections 3.4 and 3.5) and no unique, purely syntactic and algorithmic procedure for sentence interpretation (section 3.6).

I hope these comparisons may help obtain a better, clearer understanding of open compositionality, its view of semantics as decision-making, and its cognition-first methodology. The remaining chapters of this book are intended to show how this view of natural language can be fruitfully applied to explain some of the most recalcitrant phenomena considered by the philosophical tradition.

Chapter 5

Lexical Processing Architecture and Substitution Failure

The phenomenon of substitution failure is one of the most longstanding problems in philosophy of language, presenting a substantial challenge to any satisfactory account of how competent speakers use and understand proper names. There is substitution failure whenever the substitution of coreferential names fails to preserve the meaning of a given complex expression. This phenomenon is seriously problematic for the closed view because it directly entails a failure of compositionality. According to the closed view, if coreferential name substitution fails to preserve the meaning then it fails to preserve truth-conditions. If so, then the meaning of the relevant complex expression *E* is not fully determined by the meaning of its expression parts, because coreferential names are meaning-wise, and truth-conditionally, equivalent. The problem is especially acute for those within the traditional view that wish to keep an intuitive referentialist understanding of proper names according to which they contribute only their referents to the meaning of a complex expression (see Kripke 1980; Donellan 1966 and 1970; Putnam 1975; and Marcus 1961; see also Stalnaker 1978; Recanati 1993; Salmon 1986; and Soames 2002). Alternatively, instead of abandoning *strong compositionality*, a good number of philosophers have chosen to defend descriptivist theories of proper names by simply adding new semantic dimensions for proper names (on top of their reference) to fill in any information that may be needed to keep compositionality (see Frege 1892; Russell 1905; Lewis 1984; Stanley 1997; Jackson 1998; Kroon 2004; and Elbourne 2005).

Now, even though it has been commonly regarded as a problem for the closed view, substitution failure remains a serious theoretical challenge even after abandoning *strong compositionality*. It is puzzling to see that by merely using different yet coreferential proper names in the same context,

with apparently the same appropriateness and economical restrictions, speakers manage to influence the hearer into drawing different, sometimes incompatible interpretations. Why is this the case? Why is it that competent subjects consistently assume that sentences involving different names demand different interpretations, even when the names are coreferential? The phenomenon becomes more puzzling when we realize that it emerges in contexts that include enough information suggesting there should be no such difference in the interpretation. There is substitution failure even in contexts including the presupposition that the relevant names have the same interpretation. Intuitively, if they have the same interpretation, and the same appropriateness and economical restrictions apply, the sentences involving different coreferential names should at least have similar or compatible interpretations, but they do not. In this chapter I will offer an explanation of this phenomenon from the point of view of open compositionality.

I start, in section 5.1, by describing the phenomenon in need of explanation. I will argue that there is substitution failure not only when there are differences in truth-values—e.g., for reports of attitudinal mental states, or within simple sentences—but also when there are different interpretations that may be said to have the same truth-value. In sections 5.2 to 5.4, I present some psychological data on the processing of names. I focus mainly on studies on memory for proper names (section 5.2), heuristic strategies for acquiring new proper names (section 5.3), and the processing architecture for storage and retrieval of proper names (section 5.4). Section 5.5 offers an account that explains why and how speakers may process proper names differentially, regardless of whether they are indeed coreferential. In section 5.6, I put the theory to work by showing how it illuminates Frege's puzzle of informativeness as well as Kripke's puzzle about belief. Due to space limitations my discussion of these puzzles will be rough and brief.

5.1 THE PHENOMENON AND ITS SCOPE

The tradition has generally focused on how coreferential proper names fail to substitute within propositional attitude ascriptions. Consider the example in sentences in (1) and (2).

(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*.

At first glance, sentences (2a) and (2b) appear to be consistent with each other. However, it follows from (1a) to (1c) that the names "Clare Quilty" and "Dr. Zemph" are coreferential. But, if we substitute these coreferential names in either (2a) or (2b) the sentences are no longer consistent with each other. By substituting "Clare Quilty" for "Dr. Zemph" (2b) contradicts (2a). Initially it would seem as if there is only substitution failure within (2a) and (2b). The philosophical tradition has mainly followed Frege (1892) on this, given that it is only within (2a) and (2b) that substituting coreferential names results in a shift in truth-values and, *a fortiori*, in truth-conditions. It is not hard to see why the closed view would focus only on this instance of substitution failure. But once we depart from *closed compositionality* it becomes quite obvious that substitution failure abounds. In general, the substitution of coreferential names changes the interpretation of what appears to be the same sentence in the same context and with the same appropriateness and economical restrictions. This is evident in sentences (1a) and (1b). Frege (1892) famously notes that identity statements such as (1c) are specially prone to exhibiting failures of substitution, running the risk of becoming trivial if one of the names is substituted for the other. Thus, for example, if we substitute "Dr. Zemph" for "Clare Quilty" in (1c) the result is a sentence, such as (1d), so trivial that competent speakers would have trouble finding an appropriate interpretation.

(1d) Clare Quilty is Clare Quilty.

Since Frege (1892), cases of substitution failure such as (1c) and (1d) have been considered to be a separate phenomenon, giving place to what is commonly known as the problem of *informativeness*—since the substitution of coreferential names gives place to a difference in informativeness between the relevant sentences, in this case (1c) and (1d). From the point of view of *open compositionality*, specially given that semantics is a matter of decision-making (and not of truth-conditional *calculi*), the difference between (2a)—(2b) and (1c)—(1d) becomes immaterial. In both cases the substitution of coreferential names modifies the appropriate interpretation of the sentence.

Now, independently of these considerations, there are good reasons 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 allegedly demands for a different interpretation than (4b).

(4a) It is a priori that Dr. Zemph is Dr. Zemph.
(4b) It is a priori that Dr. Zemph is Clare Quilty.

Thus, it seems that there may be substitution failure wherever there is a proper name involved. To properly understand why this is the case we need to understand what is so special about proper names. To do so the cognition-first methodology asks us to first look at the cognitive psychological resources deployed for the acquisition, processing, and understanding of proper names among competent subjects. I will do so in what follows.

5.2 MEMORY FOR PROPER NAMES

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" (Kaplan, 1990, 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 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, 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 idea 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 (e.g., Recanati 1993). According to such vies, names contribute associated mental files in the form of encyclopedia entries that include descriptive information about the referent of the name. The empirical evidence undermines theories like this.

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 or 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 especially difficult to retrieve. If descriptive information is to be always associated with the name it must always offer extra entry points from which to access the name. This would make names easier to retrieve. 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—a case of aphasia associated to name retrieval failure—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 (e.g., 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 (2011).

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. Thus, we cannot expect to account for substitution failure by appealing to descriptive information differentially associated

to each member of a set of coreferential names. Studies on memory for proper names present us with a first important, albeit negative, lesson in our search for an account of substitution failure. Assuming that the (lexical) meaning of a proper name—if there is any such thing—consists of information closely (perhaps necessarily) associated with the proper name, the evidence suggests that the substitution failure phenomenon has little or nothing to do with the so-called "lexical meaning" of proper names.

5.3 SIMPLE HEURISTICS FOR PROPER NAMES

In order to acquire a language, human infants must face what Quine (1960) famously described as the problem of "radical translation." Whenever they encounter the use of a word by an adult infants must pick one among multiple different hypothesis about its meaning. The novel word could refer either to a whole object or to just a part of it. It could also refer to a property of that object or even to a relation in which that object stands. Whenever an infant first encounters the use of the English word "dog" as applied to an object X, for example, the infant must decide whether to take "dog" to refer to dog X, the kind DOG, an undetached part of dog X, the color of dog X, or to dog X whenever it is perceived (or many other alternatives, see Quine 1960).

The problem is accentuated by the fact that, even after several encounters with the use of the relevant word, the available evidence is not enough to justify selecting one among the many alternative hypotheses. The task of deciding upon the right hypothesis is difficult. Nonetheless, by the time they are 18 months of age, infants appear to have mastered the task of word acquisition. Their vocabularies consist of over a hundred words including both proper and common nouns (though not in equal proportion, see Nelson 1973 and 1974).

As Quine (1960) forcefully argues, this problem cannot be solved by carefully testing each individual hypothesis against the evidence. So there must be another way in which human infants solve the problem of "radical translation." The evidence on cognitive and lexical development (see Markman 1991, for a detailed review and discussion) suggests that infants aid themselves by honoring some defeasible yet useful heuristic assumptions known as the *whole-object* and *taxonomic* assumptions:

> (W)hen children hear a novel label, they assume that the label refers to the object as a whole, and not to its parts, substance, or color, and so on, and that it refers to other objects of the same kind or same taxonomic category, and not to objects that are thematically related. (Markman 1991, 73)

Infants also acquire words that refer not to whole objects but to properties of these. To do so infants appear to honor the assumption of mutual exclusivity:

> In addition to using the whole-object and taxonomic assumptions, children constrain word meanings by assuming at first that words are mutually exclusive—that each object has only one label. Given the nature and function of category terms, they often tend to be mutually exclusive. A single object cannot be both a chair and a dresser or a chair and a table. (Markman 1991, 83)

Markman's work shows how this is true with respect to common nouns. Other studies (see Hall 1996 and 2002) show that this is also true of proper nouns (see also Baldwin, Markman, Bill, Desjardins, Irwin, and Tidball 1996; Bloom 2000, Carey 1997; and Soja, Carey, and Spelke 1992). Competent speakers seem to assume not only that each different common noun labels a different property, but also that each different proper name labels (or refers to) a different object. This is what Hall (2002) dubs the "one name per object" assumption.

Hall (2002) reviews different studies on how semantic knowledge supports proper name acquisition. He considers different heuristic strategies, two of which are of great importance for our discussion. Speakers seem to presuppose, first, that there is only *one object per name* and, second, that there is *only one name per object*. Katz, Baker, and Macnamara (1984) first noted a third strategy, "salient individuals," in a study with three-year-olds. They noticed 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. Let me focus then on the two initial strategies for acquiring proper names: *one object per name* and *one name per object*.

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 four-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 from group one would interpret the novel word as a name, while children from 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 four-year-olds assumed that proper names could only apply to single objects. Children from group one refused to extend the use of the novel word to apply to more than one object, while children from group two did not.

In a different study, Hall (1996) also showed that the *one object per name* assumption is a default yet revisable heuristic strategy. 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 if given enough evidence to the contrary. 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 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 to 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 presents evidence supporting two different claims. First, the evidence suggests 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. Alternatively 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, competent subjects 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.

In other words, in contexts that provide a single salient referent, competent speakers normally interpret novel words as proper names *and* when interpreting a word as a name subjects 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 follow a simple heuristic strategy according to which there is *only one object per name*. But, if competent speakers presuppose that to each name corresponds only one object, it seems natural to expect them to 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 to only one name*.

Forty-eight four- and five-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 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 (1999) 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, 88)

These studies, presented in Hall (2002), suggest that, from a very early stage of development, speaker knowledge of proper names includes two default heuristic strategies according to which (in general) there is only one object per name and one name per object. According to the former, a single proper name refers to only one object—homonyms constitute a special case. According to the latter, individual objects have only one name or, better put, different names must refer to different objects. To say that both heuristic assumptions are "default" is to underscore the fact that speakers typically follow them yet, when given enough evidence to the contrary, subjects do revise their assumptions and accept exceptions.

These assumed heuristic strategies tell us a lot about what speakers normally do when acquiring different names that, by accident or not, have a common referent, which is precisely the scenario where substitution failure appears. The empirical evidence suggests that speakers start by assuming that a given name, for example *N*, refers to only one individual and that such name *N* will be enough to refer to that individual in all circumstances. Hence, competent subjects assume as well that a different name, say *M*, will refer to a different individual. It follows that cases where two different names, say *N* and *M*, refer to the same object will be considered exceptions and, hence, will be considered names for different objects unless given enough evidence to the contrary. In short, the evidence suggests that competent speakers follow the *principle of mutual exclusivity* as a heuristic strategy for acquiring and processing both common nouns and proper names.

Mutual Exclusivity: For any pair of labels "M" and "N," and any context *c*, assume that "N" and "M" label different things—e.g., different kinds, different properties, or different objects—unless *c* provides enough counterevidence against the claim that "N is not M."

If competent speakers do follow *mutual exclusivity* as a default heuristic strategy, it should not be a surprise, then, that competent speakers do not fix on the same interpretation for sentences 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. Competent speakers will *initially* assume that "*M*" and "*N*" are different names corresponding to different objects and, thus, will acquire them—i.e., will process and store them—as different lexical items. Doing so will have important consequences when it comes to retrieving such lexical items for subsequent uses, e.g., to answer questions, make assertions, etc. To see exactly how this works we must look into the details of lexical access for speech production. I will do so in the following section. So far, the evidence suggests that the phenomenon of substitution failure is related to simple heuristic strategies followed by speakers in order to simplify the acquisition and processing of proper names.

5.4 PROCESSING ARCHITECTURE FOR DISTINCT NAMES

The principle of mutual exclusivity plays a central role in the process of word acquisition. This role ends up determining distinctive and sometimes separate patterns of processing for each acquired word. The phenomenon of substitution failure could be a by-product of the patterns set by word acquisition in accordance with the processing architecture for speech production and comprehension in terms of access to the lexical items and their lemmas (i.e., associated information specifying semantics, syntax, pragmatics, conditions of proper use, etc.). To see how this might be, we must consider how speech production and comprehension is meant to work once the lexicon is shaped, at least in part, by the principle of mutual exclusivity.

First, competent speakers start by assuming that labels are mutually exclusive. For example, if a speaker acquires "M" and "N" in contexts that do not provide enough information against the claim that "M is not N" she will assume that "M" and "N" label different things or objects.

Second, once two words or two names (e.g., "M" and "N") are taken to label different things, each name is stored as a separate lexical item, and the information associated with each word is stored under separate lemmas, one per lexical item. This means, for example, that the information associated with "M" will not be directly related to the information associated with "N." In other words, if a competent speaker takes two words or two names

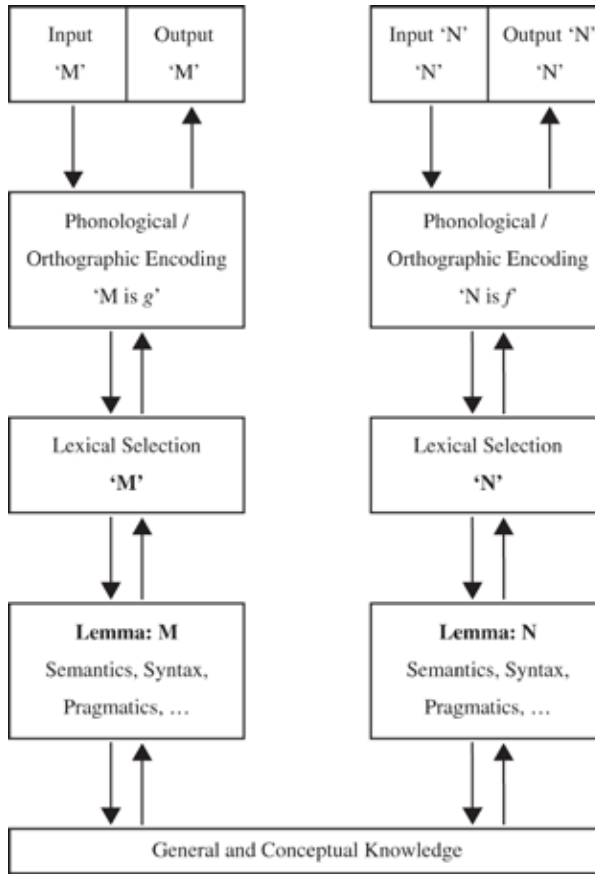

Figure 5.1 Processing architecture for distinct lexical items 'M' and 'N'.

to be mutually exclusive, the information associated with each name will be stored in such a way that all inferences involving it will not directly depend on inferences involving the information associated with the other name (or word). Figure 5.1 illustrates this point. Given that "M" and "N" are acquired as separate, distinct lexical items, they are associated with distinct lemmas. There are no arrows connecting directly the processing associated with one item with the one associated with the other, either at the lexical or at the lemma levels. The relevant words (or items) and their meanings (or lemmas) are only indirectly related, once the processing of general or conceptual knowledge is reached.

More generally, when different words are acquired as labeling different objects, the information associated to each word will be independently stored and accessed as in Figure 5.2, with no arrows directly relating one file with the other.

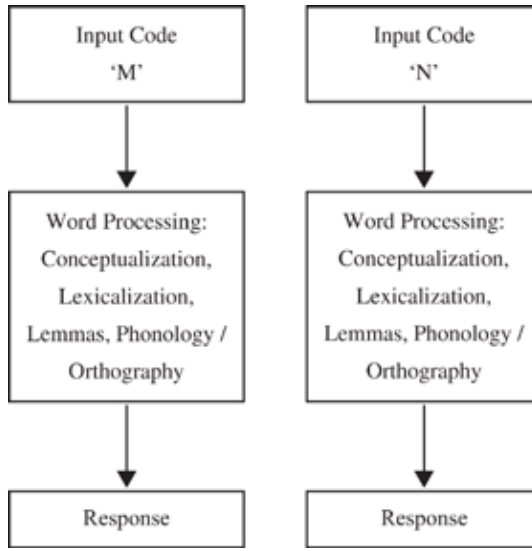

Figure 5.2 Storage and retrieval architecture for words that are taken to label different things.

Given this processing architecture, whenever two words are acquired as labeling different objects (or different properties), the corresponding lexical items and lemmas will only be indirectly related, through general knowledge and conceptual reasoning. Accessing general knowledge to identify or determine conceptual relations among different lexical items is of course, a complex cognitive task demanding conscious reflection and, thus, requiring important amounts of time and memory resources. For speakers to complete such task they must, among other things, gain access to the relevant lexical items and their lemmas.

For example, to determine the conceptual relations between "M" and "N," assuming they were acquired as different names for different objects, speakers must complete three different tasks. In some or other order, the speaker must access the lemma for each of the labels, "M" and "N." These are two distinct tasks. And, finally, upon accessing general knowledge through the associated lemmas, the speaker must reflect upon such conceptual relations and determine their nature—e.g., determine if they are the same concept or a closely associated, subordinate, independent, or distinct concept.

Comparing "M" with "N": For any competent speaker *S*, any pair of lexical items "L_1" and "L_2," *S* must:

(i) Access L_1: Access lexical item and corresponding lemma.
(ii) Access L_2: Access lexical item and corresponding lemma.
(iii) Compare: reflect upon L_1-L_2 relations.

This model directly predicts that determining which are the conceptual relations between two distinct lexical items (and their lemmas) will be a more complex, more cognitively demanding, and non-trivial task than determining the conceptual relations holding within one and the same lexical item. Identifying and determining the conceptual relations holding within one lexical item L_1 demands, at most, the completion of task (i), accessing its lemma. The resulting comparative processing difficulty, predicted by generally accepted models of speech production and comprehension, is a consequence of the *architecture* of word processing (either for production or comprehension). This means that such comparative difficulty will hold no matter which information is contained in the corresponding lemmas, in particular, it holds independently of whatever is specified by the semantics of the lexical items.

These ideas suggest a useful notion of comparative processing difficulty for distinct names:

Comparative Difficulty: for any competent speaker *S*, any pair of distinct lexical items "L_1" and "L_2," and any relation R among concepts, determining whether L_1 R L_2 is true, will be comparatively more difficult for *S* than determining whether L_1 R L_1 is, regardless of what the semantics of "L_1" and "L_2" specify.

The evidence and debate on word processing models offer a detailed account of the consequences of acquiring proper names as distinct labels for distinct objects. Regardless of whether there are in fact two distinct objects or if, for some or other reason, the subject falsely believes there are, each name will be given independently accessible lexical entries and the associated lemmas will hold no relations in between. Furthermore, if the names so acquired in fact denote one and the same object, realizing this will be a cognitively demanding task comparatively more difficult than reflecting upon a single lexical item and its corresponding lemma. After looking at studies on memory for proper names, heuristic strategies for acquiring and learning proper names, and architectural models for word processing we now have the ingredients of a satisfactory account of substitution failure.

5.5 A PROCESSING APPROACH TO SUBSTITUTION FAILURE

The studies on memory (see section 5.2) 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 anomic 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 and retrieving descriptive information (see García-Ramírez and Shatz 2011 for a detailed discussion).

Given this cognitive limitation, the studies on the heuristic strategies followed by competent subjects in their acquisition of names (see section 5.3) suggest that whenever a speaker acquires distinct novel names, she will assume by default that they belong to distinct individuals. Finally, the studies on lexical access and word processing generally agree with processing models according to which words that are acquired as denoting distinct objects are stored in unrelated files, thereby granting independent patterns of inference. When this happens, whether the names in fact are coreferential will be immaterial to the way the subject stores and retrieves them from memory. The processing patterns will be independent to the point that processing both names (*qua* lexical items) and drawing any relations in between will be comparatively more difficult, more cognitively demanding, than processing only one of them.

These results suggest an account of substitution failure based on the processing architecture above described. Recall the initial formulation of the phenomenon from section 5.1. Substitution failure takes place whenever the substitution of coreferential names induces different interpretations of what appears to be the same sentence in the same context and with the same appropriateness and economical requirements. For example, (1a) and (1b) seem to have different meanings, and so do the pairs (1c) - (1d) and (2a)—(2b)

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

(1c) Dr. Zemph is Clare Quilty.
(1d) Clare Quilty is Clare Quilty.

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

More generally, it seems that whichever sentence may include a proper name will be prone to exhibit substitution failure. The evidence I have considered suggests that this is a consequence of the special processing architecture and heuristic strategies dedicated to the acquisition, storage and retrieval of proper names.

Accessing and retrieving a name is a matter of accessing and retrieving very little information, perhaps some visual representation, hence their

retrieval difficulty. Consequently, there is little information guiding the interpretation of a proper name. To cope with this lack of information competent subjects follow a set of generally successful and cognitively effortless heuristic strategies. In the absence of evidence to the contrary, by following these heuristic strategies competent subjects will acquire distinct names (e.g., "Clare Quilty" and "Dr. Zemph") as belonging to distinct objects and, hence, assigning a different lexical entry for each name. In doing so different processing trajectories will be guaranteed for each name, regardless of which object they in fact refer to. Thus, according to this account, in order to understand (1a) and (1b) competent subjects most access distinct lexical files and, thus, follow distinct processing trajectories. It is in virtue of the latter that these sentences exhibit substitution failure according to this view. Briefly put, substituting coreferential names within (1a) and (1b) would entail substituting different, non-equivalent, processing trajectories for each sentence; hence the failure to preserve the same interpretation after the substitution takes place.

Sentences (1c) and (1d) demand a similar explanation, yet the case deserves further discussion given the place that identity statements occupy within the philosophical tradition since Frege (1892); I will do so in the following section. Now, traditionally attitude ascriptions such, (2a) and (2b), are considered to involve an extra element—e.g., modes of presentation, reference to senses, etc. Unlike (1a) and (1b), the shift in interpretation induced by the substitution of coreferential names in (2a) and (2b) also entails a shift in the sentence's truth-values. Yet, on the processing architecture theory there is no need of extra elements. To account for how the substitution of coreferential names may shift the truth-value of certain attitude ascriptions we need only look at the specific context in which such ascriptions take place and search for the most appropriate and economical interpretation within such contexts (see section 4.4 and 4.6).

It is well known that attitude ascriptions have at least two interpretations, *de re* and *de dicto*. *De re* interpretations are commonly *not* subject to substitution failure. To illustrate consider a context with no substitution failure such as C1:

> C1:
> Suppose we are talking about the history of psychoanalysis, on which Professor Humbert is an expert. He is not part of 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, and we know, of course, that "Dr. Zemph" is Clare Quilty's pseudonym. We are not sure about the merits of his *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."

C1 is a context suitable for a *de re* reading of the attitude ascription, since the focus of this conversation is what Professor Humbert thinks *of* Clare Quilty himself. An appropriate interpretation of what you said would go as follows: "Humbert thinks *of* Clare Quilty that he is. . . ." And there is no relevant difference between this interpretation and the following alternative: "Humbert thinks *of* Dr. Zemph that he is. . . ." Thus, in a context such as C1 whether the subject of the attitude ascription—i.e., Professor Humber—knows or not that the names in question are coreferential is not relevant to the interpretation. Now consider a context exhibiting substitution failure.

> C2:
> 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 Professor Humbert does not. After a while, you ask: "Did you know that Clare Quilty will be visiting us later today?" Professor Humbert is surprised: "I thought it would be Dr. Zemph presenting his new book!" You do not 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: "Professor Humbert does not believe that Clare Quilty wrote *Psychoanalysis*."

This latter utterance of mine is not appropriately interpreted as a *de re* ascription but as a *de dicto* one. 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 C1 and C2. In C1 we are discussing the history of psychoanalysis, not Professor Humbert's thoughts or behavior. Hence, the purpose of your ascription in C1 is not to explain Professor Humbert's thoughts or behavior. Rather, your attitude ascription is concerned *de re* with Clare Quilty himself and his role in the history of psychoanalysis. That is why you can be appropriately interpreted as saying "Humbert thinks *of* Clare Quilty that he. . . ." However, once the subject of ascription himself and his thoughts are at issue in the conversation (as in C2) and, most importantly, once it becomes a goal of the conversation to explain the subject's thoughts and behavior, the belief report is not concerned anymore with Clare Quilty himself.

So what would be an appropriate interpretation of my utterance in C2? There is, I think, a rather obvious and intuitive interpretation. What I am saying by uttering "Professor Humbert does not believe that Clare Quilty wrote *Psychoanalysis*" is *not* 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. I am saying, rather, something about Professor Humbert's understanding of the name

"Clare Quilty": e.g., that he does not associate "Clare Quilty" with authoring *Psychoanalysis*. After all, as Geurts (1998) observes, we ordinarily use names in order to mention them.[1]

So attitude ascriptions such as (2a) and (2b) exhibit substitution failure whenever they are used in contexts such as C2, that is, contexts where the focus is on what is said about the mental states of the subject of ascription and not on the objects that such mental states may be directed towards. When appropriately interpreted (2a) and (2b) are about Professor Humbert's failure to understand that Clare Quilty and Dr. Zemph are one and the same individual.

This brings us back to the processing architecture account. According to this account whenever two names are acquired and processed as denoting distinct objects, they will be stored differentially, with independent access and processing trajectories. If the interpreter is in such a scenario, there may be substitution failure with respect to the interpretation of *simple* sentences involving them. When it comes to reporting the mental states of a subject who happens to be in such a situation—i.e., who has acquired and processed different names as if they had different referents when in fact they are coreferential—competent speakers may choose to use attitude ascriptions involving substitution failure to report such thoughts. This explains why, in the case of (2a) and (2b), the substitution of coreferential names would intuitively imply a shift in truth-values.

It is important to note that the proposed theory—i.e., the lexical processing architecture account of substitution failure—is easily applicable to account for any instance of substitution failure, even those involving other expressions beyond proper names (e.g., demonstratives). On this view, substitution failure is owed to the existence of distinct processing patterns for independent lexical entries. It is consistent with this theory that almost any denoting expression, including proper names as well as demonstratives and perhaps also indexical expressions (for a classical example see Perry 1979; see also Perry 1993) may be used in a certain context giving place to substitution failure. If in a particular context C the relevant expressions e_1 and e_2 have independent lexical access and processing architecture, their substitution within a given sentence S may fail to preserve that sentence's interpretation in C. It has been argued that even common nouns may give place to the same phenomenon (see Kalderon 2004). The processing account here proposed exhibits a desirable degree of generality.

I would like to conclude by offering further evidence of the theory's explanatory power. In the following section I will show how the theory can account for two longstanding philosophical problems, Frege's puzzle of informativeness and Kripke's puzzle about belief.

5.6 FREGE'S AND KRIPKE'S PUZZLES

Let me now go back to the case of identity statements (1c) and (1d). These statements appear to differ substantially, even though they merely differ in the substitution of coreferential the names "Dr. Zemph" and "Clare Quilty."

(1c) Dr. Zemph is Clare Quilty.
(1d) Clare Quilty is Clare Quilty.

Frege (1892) famously argues that true identity statements involving proper names such as (1c) present us with a theoretical challenge. On the one hand, true identity statements may be informative. In certain cases, competent speakers may wish to use an identity statement to express a substantial discovery. Hence, true identity statements involving proper names do *not* merely state the trivial truth that a given object *o* is self-identical. On the other hand, if true, the names involved in an identity statement are coreferential and so may be substituted amongst each other without changing their denotation and, *a fortiori*, without changing the interpretation of the sentence itself. Yet if we substitute coreferential names within a true identity statement the result can no longer be an informative statement, it will rather be a trivial one—e.g., "Clare Quilty is Clare Quilty." Thus, the challenge posed by true identity statements is that of explaining how they may be informative—sometimes substantially so—even though they only involve coreferential proper names.

The processing architecture account here offered delivers an explanation of this puzzling phenomenon by following the principles of *open compositionality*. If sentence meaning is determined as an appropriate and economical interpretation of a given sentence in a particular situation and with certain limited information, then to determine how a sentence like "Dr. Zemph is Clare Quilty" may be informative we must look at how competent subjects may process such a sentence in contexts where it may be informative. Here is one such example, taken from Nabokov's script for Kubrick's film *Lolita*. Professor Humbert (H) is Lolita's (L) stepfather. Towards the end of the film, H interviews L about their common past.

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?
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.

H has personally met with Dr. Zemph as well as with Clare Quilty. He knows them both by acquaintance. Still, H is unable to properly associate the information he is given by L. He remembers pretty well who called him by phone, 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, H is still clueless. The end of the dialogue confirms this by pointing out, among other things, that H takes the identity sentence "Dr. Zemph is Clare Quilty" to be of great informative value.

How is this possible? H, a competent subject, is personally acquainted with the referent of both names, "Dr. Zemph" and "Clare Quilty." Following the heuristic principles for acquiring and processing proper names, according to which there is one name per object and one object per name, H has acquired these names as if they were names of different objects—i.e., storing them in independently accessible files with independent processing trajectories. H has not been given enough evidence against this hypothesis, so it is plausible to think (see sections 5.3 and 5.4) that H takes them to correspond to different objects. Because of this H has not been able to realize that "Dr. Zemph" and "Clare Quilty" *do not* denote different objects. This would explain why H remains clueless all through the conversation with L, until he is explicitly told that Dr. Zemph is Clare Quilty.

L knows that "Dr. Zemph" and "Clare Quilty" are different names of the same individual, but she realizes that H assumes that they are names of different people. This knowledge of H's understanding (or lack thereof) becomes salient in the context of the conversation. L can exploit it in order to convey exactly that information which would put an end to H's ignorance. The intuitive interpretation of L's utterance of "Dr. Zemph is Clare Quilty" is *not* that "the man at the motel is the man in the car." L has already presented this kind of information and H has failed to draw the right inferences. The intuitive interpretation is, rather, something like "Dr. Zemph and Clare Quilty are not two different people."

True identity statements involving different names—e.g., "Dr. Zemph is Clare Quilty"—may be substantially informative to competent subjects who justifiably assume that the relevant names belong to different objects. For

those subjects, coming to learn the truth of the identity statement is the best way to rectify their assumptions. It is these subjects who will find an appropriate an economical interpretation of an identity statement of the form "A is B," such as (1c), to be highly informative, even if the names involved do not differ denotation-wise from those used in the corresponding but trivial identity statement of the form "A is A," such as (1d).

We can now meet Frege's challenge. Identity statements such as (1c) are informative, even though they only involve coreferential proper names, because they can help rectify a competent subject's ignorance about those names' coreferentiality. This ignorance, it is important to note, is compatible with the subject's knowledge of what each name refers to. So long as she lacks enough evidence to the contrary, the subject will assume that to each name there corresponds a different object, and so will acquire each name as corresponding to an independent lexical file. In a situation such as this, the identity statements (1c) and (1d) are not equivalent to each other, neither in terms of processing needed nor in terms of the inferences H may draw from them. In terms of processing, to properly understand (1d) H must merely access a single lexical item and its lemma (see section 5.4). This is not the case with (1c). Since it involves distinct lexical items, H must access a second lexical item and its lemma. H may then reflect as to whether the concepts associated with each lemma are in fact about one and the same referent, allowing H to obtain important information, ultimately helping him rectify his ignorance.

To conclude this chapter let me consider Kripke's (1979) puzzle about belief and how the proposed theory may shed some light on it. According to Kripke (1979), there's something puzzling about belief, for competent subjects may form incompatible beliefs as a result of assuming that a single proper name, which even according to the heuristics of proper names mentioned in this chapter (see section 5.3) *should be* understood as naming only one object, is understood as naming two different things. How is this possible? Consider Kripke's own example:

> 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:
>
> (5a) 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
>
> (5b) Peter believes that Paderewski had no musical talent.
>
> (Kripke 1979, 265)

How can anyone be in Peter's situation? We can find a simple answer in the processing architecture account, more specifically by looking into the heuristics of proper name acquisition and processing. In section 5.3 I offered evidence suggesting that in general speakers assume that to every name corresponds only one referent. This suggests that Peter should always treat "Paderewski" as denoting a single object that is both a musician and a politician. Peter, however, does exactly the opposite. Yet, this does not falsify the processing account. As I said in 5.3 and throughout, heuristic strategies are quick, simple, and flexible guides meant to help a subject adapt to the specific task at hand. Heuristic strategies are not always successful, and that explains why they are revisable. If there is enough evidence to the contrary, speakers may actually decide against what the heuristic strategy mandates. With enough evidence in hand, competent speakers may conclude that what would otherwise appear to be a single name are in fact two distinct homophonic names denoting distinct objects and, hence, acquire them as having independent lexical access and processing trajectories, thus allowing for beliefs about one that are incompatible with respect to those associated with the other. Kripke's Peter appears to be in a situation just like this.

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 individual. But he has also gained access to extra information. He is first introduced to a use of "Paderewski" that refers to a famous pianist, and later to a different use of it that refers to a politician. Peter is convinced that politicians are terrible musicians. For Peter, this is enough evidence to assume 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* proper names. The theory I have proposed predicts that in these cases subjects will understand such uses of "Paderewski" as evidence of distinct proper names with distinct denotations, and that this will carry important differences in lexical processing architecture (see section 5.4).

This explains why Peter may be said to believe both that Paderewski had musical talent and that Paderewski lacked musical talent without ascribing him irrationality or any kind of puzzling mental representational state. There is nothing irrational, or puzzling, about assigning incompatible properties to *different* objects. And there is also nothing irrational, or puzzling, in falsely believing that what appears to be a single name, with a single referent, are in fact a pair of homophonic names with distinct reference. It is the conjunction of both that results in the peculiar situation of Peter. The processing architecture account directly explains the latter by looking into the heuristics of

proper name acquisition and processing. The former is made possible by the latter.

I believe the lexical processing architecture approach to substitution failure constitutes an outstanding theory. It explains all the relevant phenomena, including possible extensions of it involving other denoting singular expressions and even common nouns. The account is also ontologically parsimonious, especially when compared with alternative philosophical accounts that end up postulating dubious entities such as senses, modes of presentation, or propositional guises (see Frege 1892; Recanati 2012b; Salmon 1986; and Chalmers 2002, among others). It relies on processing models for lexical items, but such models may be had for free as they are needed for independent purposes—e.g., to account for actual language processing. The account is a product of the *cognition-first* methodology endorsed by open compositionality. As such, it also observes the *Lewisian Compromise*. Having these theoretical virtues the account is worth serious consideration and an important witness of the theoretical success of open compositionality. In the following chapter I will discuss one more theoretical product of open compositionality, this time concerning the use of empty names in linguistic practice.

NOTE

1. According to Geurts (1998), "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, 291)

Chapter 6

Decoupled Representations and Empty Names

How do competent speakers use names without referents (e.g., "Pegasus," "Hamlet," "Vulcan," etc.) to convey meaningful, truth-evaluable and full-blown informational contents that may even guide their behavior?[1] In this chapter, I want to offer an account of empty names that does not take them to have descriptive semantic content. Doing so will allow us to keep a homogenous and empirically supported (see García-Ramírez and Shatz 2011) theory of meaning for referential uses of names (empty or otherwise). Following the *cognition-first* methodology endorsed by open compositionality the theory purports to be cognitive, as opposed to semantic or even linguistic. I will begin in section 6.1 by describing the psychological context within which this theory is conceived and the philosophical advantages of the empirical account here presented. Section 6.2 presents the cognitive theory of empty names. Section 6.3 describes the phenomena that, according to the literature, any theory of empty names must explain and how the theory I propose accounts for it. I conclude by presenting the virtues of this theory in section 6.4.

6.1 PRETENSE AND IMAGININGS

When a four-year-old pretends a banana is a telephone, she does not thereby believe that the banana is a telephone or that bananas are telephones. When she pretends of an empty cup that it is full, she behaves as if the contents were spilled when it is turned upside down. For this to be the case, children must have a way to *quarantine* their pretense from their beliefs, a way to distinguish between a referent and a prop, and a way to develop initial acts of pretense into complex games of make-believe.

Generally speaking, there are two competing accounts of pretense available in the literature: behaviorism (see Lillard 1993 and 2001; and Nichols and Stich 2000 and 2003) and mentalism (see Leslie 1987, 1994, and 2002; 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. 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, I will focus on Leslie's decoupling mechanism. Aside from this, I will remain neutral between behaviorists and mentalists about pretense.

To account for the development of games of make-believe, the decoupling account offers a mechanism (i.e., the decoupler) that produces the representations needed, in some cases by using objects as props. According to Leslie (1987), the referential, truth-conditional, and existential commitments of the pretended representations are also quarantined. Thus, another goal of the decoupler (see Leslie 1987 and 1994) is to divorce the representation from its content in order to, later on, produce a representation by assigning an interpretation to the decoupled representation.

The decoupler has three functions to which three kinds of mental representations correspond. It is important, then, to distinguish between the latter three. First, there are representations which purport to adequately represent the subject's environment according to the information received through perception, these are "primary representations." Second, there are representations which can be viewed as mental analogues of quoted sentences, they result from divorcing primary representations from their referential, truth, and existential conditions; call these "decoupled representations." Third, there are representations that result from interpreting decoupled representations. These tertiary representations offer proper objects of propositional attitudes that may adequately represent psychological states.

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 6.1 offers a view of the mechanism itself. Solid arrows correspond to input / output relations. 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).

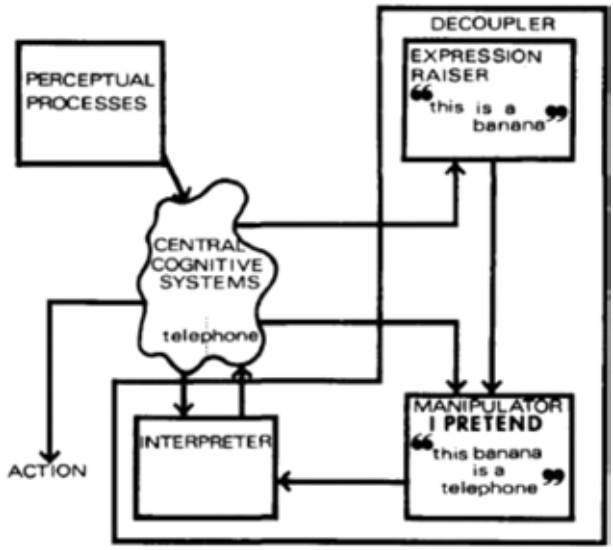

Figure 6.1 The decoupling mechanism (taken from Leslie, 1987).

As I said before, the decoupler has three different functions. The first function or subprocess consists in copying the primary representation and divorcing this copy from its semantic values. One 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. This does not imply that their ordinary semantic content of the original representation is lost. Such information may remain in central memory and may be used later on in the process of interpretation—the decoupling process has access to central memory (see figure 6.1). If, for example, the subject has perceptually acquired the representation Φ: *the cup in front of me is full*, she can then use a copy of Φ, say Φ* as part of a game of make-believe. To do so she must first decouple Φ* from its referential, truth, and existential conditions. The result will be a decoupled representation such as Φ*: "*the cup in front of me is full.*" Since Φ* is divorced from its ordinary semantic content, it prompts the question: how does one go on interpreting it for the purposes of the pretense? The remaining two subprocesses, manipulation and interpretation, take care of this.

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. Perhaps

of more general philosophical interest is the fact that this explains how decoupled representations can still figure in inferential processes. Decoupled representations can still play a role in the relevant inferential processes, once they are interpreted. The subjects may associate Φ* with the rule ρ: *if you turn a full cup upside down its contents are spilled*. Once Φ* receives an interpretation, this will allow the subject to infer Ψ*: "*there is water over the table"* if she sees that the relevant cup is turned upside down. This is possible given that the process of interpretation receives and delivers information from and to central memory (see Figure 6.1).

The third and last subprocess is that of interpretation, which consists in the assignment of the relevant reference, truth, and existential conditions to the decoupled representation in order to deliver a proper object of an attitude to central cognitive systems. To do so, the decoupler makes use of any available object (e.g., perceived objects, images, or any other available through central memory) to determine the referential conditions; sometimes this requires the decoupler to use the object as a prop. Since the process of interpretation has access to central memory (see Figure 6.1) and, thus, to information gained through perception, the process may arbitrarily assign perceived objects as referents of the corresponding representation.

The decoupler is a mechanism that involves several cognitive processes the goal of which is to deliver mental representations. It is not itself a mental representation—e.g., an item in the language of thought, a thought vehicle or anything of that sort. It does involve the use of mental representations, one of which will be central for my account, those that I call "cognitive depictions" (CDs). I will say more on CDs and how they relate to other mental representations proposed in the literature (e.g., items in the language of thought) in section 6.2.

As I will show later on, the decoupler mechanism is meant to be perceptually triggered. As such, it does not distinguish between genuine empty names that are not known to be empty and non-empty names that are also not known to be non-empty. I explain why this is the case and why it is not a problem for my theory throughout section 6.2. The decoupler has several advantageous characteristics, all of which are of substantial philosophical interest. It is cognitively independent, attitudinally general, interpretatively autonomous, and flexible.

Cognitive independence: the decoupler is an independent process, separate from general cognition and available for different tasks.

Attitude general: the decoupler delivers representations that are available for distinct types of mental representational states: i.e., belief, desire, pretense, etc.

Interpretative autonomy: the decoupler freely determines which are to be the referential, truth, and existential conditions of the decoupled representations it interprets.

Flexibility: as a consequence, when the referential conditions are fixed by a prop, the decoupler is flexible enough to change the object that is meant to be the reference-fixing prop without thereby changing the representation.

Its cognitive independence allows the decoupler to work even when other representation processing mechanisms (e.g., language processing) are not involved. Its attitudinal generality explains why subjects may sometimes seriously believe the representations that are produced by the decoupler, representations that seem to be more like imaginings. This, in turn, explains why subjects may guide their behavior based upon such representations. Thanks to its interpretative autonomy, the decoupler may interpret and, thus, make meaningful and understandable, representations that, say, syntactically guided processing mechanisms may find uninterpretable. In this sense, the decoupler may be invulnerable to limitations (e.g., semantic) applying to purely syntactic/semantic mechanisms. Finally, its being flexible allows the decoupler to maintain representational consistency when there is no single unique object that may do the referent's job: if there is no referent, different objects may do so long as they are props and, hence, not the referent but representations of it. But flexibility is not necessity: the decoupler *may* but *need not* enforce such flexible interpreting. In each case, the context will determine whether the interpretation is to be as flexible as described or not.

I believe all these features make of the decoupler an ideal candidate for the mechanism that underlies a normal competent subject's understanding of empty names and other referential representations without referents.

6.2 THE COGNITIVE ACCOUNT OF EMPTY NAMES

The basic tenet of the cognitive theory is that competent speakers interpret empty names by means of the decoupling mechanism just described. More specifically the claim is that a specific kind of representation produced by this mechanism, what I call "cognitive depictions" (CDs), is a good candidate for the role of content of thoughts and assertions involving empty names. Perception plays a central role in this account: it triggers the decoupling mechanism. Thus, in order to determine whether the mechanism is used to interpret a given referential term in a given context, we need to determine whether there is a perceptually available salient object that may be used as the referent of the term in that context. Before getting into this let me say with some detail what a CD is.

I hesitate to identify CDs with some or other representational notion in the literature. I hope it will be enough to give their job description while comparing them to other already available notions: propositions, items in the language of thought (ILTs), and thought vehicles.

Like propositions, CDs can represent possibilities. Unlike propositions, CDs need not represent possibilities. A CD can also depict representations of possibilities. Just like a painting of a squared circle need not depict a possibility (since squared circles are impossible) but merely a representation of a possibility, so can CDs. The explanation is rather simple. Propositions (under most interpretations at least) are made up of logical space; CDs are made up of manipulated, decoupled, and interpreted perceptual representations. They are, most importantly, the product of the decoupling mechanism.

In virtue of this, CDs allow subjects to represent what propositions cannot. 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, as it requires that we take a single object (i.e., "The Book of Sand") to be both finite and infinite—for further examples see Borges, *The Aleph*. Insofar 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. Similar considerations apply to other cases. We can, for example, understand what it is for something to be a regular geometrical body made up of regular heptagons. Yet, we all know it is not possible for there

to be a regular geometrical body made up of figures that have more than five equal sides.

An interpretation is not a set of sentences, descriptions, or images that determine the meaning or semantics of a representation or a name. According to the description I gave of the decoupling mechanism, an interpretation of a CD is better understood as an arbitrary assignment of referential, truth, and existential conditions to a decoupled and manipulated representation already available from central memory (see section 6.1). Such an assignment can be as arbitrary as needed, depending on the context. A subject can gain, through perception, a representation of a finite book-like object and may also gain (perhaps not through perception) a representation of what it is for something to be infinite. The decoupler may retrieve such representations, separate them from their ordinary referential, truth, and existential conditions and reinterpret them. The result will be a CD depicting an infinite book-like object that one can look at and hold in one's hand, just like Borges describes it. It is important to note that the cognitive significance of CDs is not to represent possible truths, but to constitute objects or, if you like, vehicles of thought. Thus, with respect to CDs one need not ask how one ought to represent contradictions as possible truths. The question CDs are meant to answer is this: how does one make sense of Borges' text? The answer is simple: by forming a corresponding CD in the way just described. Similarly, CDs answer the question: how does one make sense of representations using empty names?

So CDs are closer to being thought vehicles (i.e., mental representations that can express propositions) than to being propositions. Are they just like items in the language of thought? That depends on how flexible the latter can be.

Consider now a theatrical representation. Peter and Paul observe two different performances of Shakespeare's *Hamlet* at the National Theatre. They have never seen, read, or heard about Shakespeare's work before. 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 φ: *that Hamlet is courageous*. It seems uncontroversial to say that both, Peter and Paul, share a common representational mental state directed towards φ even though they both use different props (i.e., Day-Lewis and Charleson) as stand-ins for Hamlet.

This partly illustrates the flexibility of the decoupler mechanism and of the CDs it produces. Take the decoupled representation "Hamlet is courageous." The decoupler may interpret it by assigning Day-Lewis as the referential condition of "Hamlet," but it may also assign Charleson. Either way, the resulting CD is the same for the purposes of the pretense in the context of the theatrical representation. CDs are flexible with respect to their referential

conditions: having different referential conditions is not enough for there to be different CDs.

The same happens the other way around. There can be different CDs even if they all have one and the same referential condition. Consider, for example, another theatrical representation. This time it is Sophocles' *Oedipus* and a single virtuous actress as the stand-in for all three Laius, Jocasta, and Oedipus. To understand this representation one must pretend that *Oedipus killed Laius and married Jocasta* and not that *Laius killed Oedipus and married Jocasta*. And one may do so even if one knows that one must assign the same virtuous actress as the referential condition for all "Laius," "Yocasta," and "Oedipus." CDs are flexible with respect to their referential conditions in this other sense then: having the same referential condition is not enough for there to be a single CD.

With this flexibility in mind we can ask: can CDs be items in the language of thought? According to the language of thought hypothesis, mental representations are formulas of a language-like representational system; they may be combined syntactically and semantically; they are causally efficacious mainly in virtue of their syntactic properties; and they provide the human mind with a domain for mental processing. If this is all there is to an item in the language of thought, then we can accept that a CD is an item in the language of thought.

Some (see Fodor 2001 and 2008) believe that, on top of this, the language of thought (LOT) must be compositional: the content of any complex formula of LOT is fully determined by (inherited from) the content of its constituent parts and the logical (or syntactic) form of the complex formula. If this is correct and formulas of LOT must observe compositionality, then CDs cannot be elements of LOT. CDs with identical referential conditions may nevertheless fail to substitute, and failures of substitution (see chapter 5) are failures of compositionality.

It is not clear which version of compositionality Fodor has in mind. *Open compositionality*, and with it the account of empty names here proposed, opposes only *strong compositionality* (see sections 2.1 and 2.2). Throughout this book (see specially chapter 3), I have presented multiple reasons to think that natural languages are not compositionally closed and, thus, fail to observe *strong compositionality*. The empirical evidence on language acquisition and development shows, briefly put, that humans do not have the cognitive means to acquire and develop a language observing *strong compositionality*. These same reasons apply to the alleged compositionality of LOT. It would be very surprising if humans were to have the resources for a LOT observing *strong compositionality* and still a lack of them for natural languages since LOT is, by hypothesis, closely (and meaning-wise) tied to natural language. Aside from this, there is one final reason to think that LOT

need not be compositionally closed. According to Fodor (2001), the compositionality of LOT is needed to account for the productivity and sistematicity of human thought. But neither productivity nor sistematicity (or computability) require *strong compositionality*, as they are compatible with much weaker versions (see sections 2.1 and 2.2). Furthermore, a satisfactory and empirically motivated account of productivity and sistematicity is available from the standpoint of *open compositionality* (see section 4.5). So there is no reason to think that CDs may not be items in the LOT.

On this view, we can understand CDs as thought vehicles that are part of a (compositionally open) language of thought. As such, a CD may be said to be inside the head. It is by means of a CD that a subject may have an attitude towards certain content (whatever is fixed during the process of interpretation). This gives us an idea of what a CD is, but there is more to be said: how are CDs related to perception? Do they depend on the absence of actual referents?

CDs are a product of the decoupler mechanism. The decoupler, in turn, is perceptually triggered: whether CDs are invoked to interpret a given referential term depends on: (i) whether there is a perceptually available salient object; and (ii) whether there is a pretense involved in the use of the referential term. If there is no perceptually available salient object, then the decoupler will be triggered and a CD will be produced to understand such representation. If there is some such object but also the pretense that such object be a stand-in or a prop, then the decoupler will also be triggered.

To clarify consider the three different contexts in which the decoupler may be triggered to aid a subject in the process of acquisition of a referential term.

C_1: There is a perceptually available salient object, but it is part of an ongoing pretense.
C_2: There is no perceptually available salient object and there is an ongoing pretense.
C_3: There is no perceptually available salient object and there is no ongoing pretense.

C_1 to C_3 are contexts in which speakers may be asked to interpret empty referential expressions. C_1 is a case where the salient object is being used as a prop: e.g., when speakers use "Santa" to refer to a real person that pretends to be Santa. In such a case the subject deals with an empty name in virtue of the fact that the salient object is not the referent of the name according to the context itself.

Contexts such as C_2 are similar to those like C_1 in that speakers pretend that there is a referent. In C_2 cases, however, there is no perceptually available salient object. This usually happens, for example, when subjects interpret a

fictional text that makes use of empty names. In both C_1 and C_2 cases the decoupler helps by assigning adequate referential conditions. This is possible partly in virtue of the decoupler's interpretative autonomy and flexibility. The decoupler freely interprets the empty name (e.g., "Hamlet") by assigning an object that is well suited for the purpose, sometimes this might be a perceptually available object that is used as a prop (in C_1) or even an image retrieved from memory—perhaps this happens in some C_2 cases.

C_3 cases are rather distinct: there is no perceptually available salient object but there is also no pretense. On this view, there is no cognitive difference between empty names that are not known to be empty and genuine non-empty names that are not known to be non-empty. For example, consider Le Verrier's famous case. Back in the nineteenth century Le Verrier observed disturbances in the orbits of Mercury and Uranus that were not predicted by the currently accepted laws. To explain away this trouble Le Verrier postulated that two different planets were responsible for this: Vulcan for Mercury and Neptune for Uranus. At the time of the postulation Le Verrier had no more evidence of the existence of Neptune than he had of the existence of Vulcan. Later research confirmed the existence of Neptune, and proved there was no Vulcan. There is no difference between "Vulcan" and "Neptune" at the time of Le Verrier's postulation. In both cases the decoupler is triggered, since at the time there was no perceptually available salient object. Later on subjects may find that the relevant names do have a referent and become perceptually aware of them, if so then the interpretation of the relevant CD changes. Such is the case of "Neptune" after the planet was sighted. The subject presumably reinterprets "Neptune" by retrieving it from central memory, decoupling it from its previous referential conditions and reassigning it a referent, this time the planet, which becomes available through perception. I said CDs were vehicles of thought that could express propositions. Cases like "Neptune" are such that the CD in fact expresses a proposition.

Thus, the existence of an actual referent is consistent with the claim that a given referential term is interpreted by means of a CD. As with "Neptune" there can be CDs associated with referential terms that have an actual referent. A difference may arise when the actual referent also happens to be perceptually available, but even then it depends on whether the context includes some form of pretense or not. CDs will not be associated with referential terms, on this view, whenever there is a perceptually salient object and no pretense involved.

It is plausible to assume that the decoupler is the relevant mechanism behind the interpretation of referential terms as they appear in cases C_1 to C_3. Now, meaningful uses of empty names take place in either one of these contexts. My claim is that competent speakers find such empty names meaningful, whether they take them to be empty (as in C_1 and C_2) or not (as in C_3),

in virtue of the decoupler mechanism: i.e., because they associate such empty names with some or other CD. In cases where the name is taken to be empty, it will be meaningful in virtue of the assignment of a prop as the referential condition of the relevant CD. Thus, it is not because speakers (semantically or pragmatically) associate descriptive information with the relevant empty name that they take such names to be meaningful.

This, of course, does not mean that the prop is determined to be the semantic content of the empty name. The theory I defend is not a theory of the semantics of empty names, but an account of the intelligibility of empty names: it claims that such intelligibility is mainly owed to a cognitive (i.e., non-linguistic) process of interpretation. In this sense, the cognitive theory is compatible with a view that takes empty names to have no semantic content. Naturally, it is also compatible with a view that takes all non-empty names to have a semantic content, even those that are not known to be non-empty. I will say more on the semantics/cognition relation below.

I hope this is enough to characterize cognitive depictions and their role in the interpretation of empty names. It is important to note that the distinction between CDs and propositions (and between their corresponding attitudes) is not meant to be consciously available to the subject. I do not take speakers to be aware of or to know that they ascribe different kinds of attitudes depending on whether they use empty names or not. Let me now say a few words on the attitudes that are associated with CDs.

If there may be CDs associated to genuine non-empty names then the difference between what might be called a "CD attitude" and, say, a propositional attitude, does not depend on whether there is an actual referent or not. It depends on whether the subject is perceptually aware of the referent or not. If the name is non-empty and the subject is perceptually aware of the referent, then there is no reason to take the subject to have a CD attitude (see "Neptune" case above).

Yet, what might be called "CD-attitudes" are importantly similar to propositional attitudes. There is, first, an important psychological similarity. Even though they differ in virtue of being directed towards different kinds of objects, they are both guided by their representational content. If I believe that Santa is coming—i.e., if my belief state is directed to the relevant CD—and I believe that Santa loves cookies, I will make sure to leave cookies by the tree. Thus, CD and propositional attitudes cause behavior in exactly the same way.

Second, there is also an important epistemological similarity. Both CD 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 such as belief—see Platts (1979), Anscombe (1957), Humberstone (1992), and Velleman (1992)—it will be adequate (or true) depending on what goes

on in the actual world and regardless of whether the content is a proposition or a CD. If Le Verrier believed that Vulcan caused perturbances in the orbit of Mercury, his belief will be said to be epistemically inadequate even if it is a CD attitude. Given that the decoupler is attitude-general (see section 1.1), not only pretense-like attitudes can be directed towards a CD.

As mental representations, the relevant tokens and types of CDs may be identified 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). As such, there is a set of criteria associated to the subprocesses of decoupling (see section 6.1) that may help identify a given CD depending on its associated: decoupled representation, patterns of inference, and referential, truth, and existential conditions. A further criterion for identifying a CD is given by its causal history. If two CDs among different subjects or within the same one are part of a common history that relates them causally, they may be taken to be the same representation.

It seems plausible to think that the representations the decoupler produces may not be purely semantic. So I find it useful to characterize such representations not in terms of their truth-conditions or the semantic contribution of their parts, but in terms of the sentences they satisfy. So, for example, the proposition that *Jon is coming tonight* (whatever that may be) is said to satisfy the sentence in (1).

(1) Jon is coming tonight.

Now, given that "Holmes" is an empty name, there seems to be nothing that satisfies the sentence in 2—but see Sainsbury (2005).

(2) Holmes is a detective.

The theory I propose claims the decoupler is involved in the interpretation of sentences such as (2) and that it may do so by using props. The use of props requires that we distinguish between possibilities and representations of possibilities. For example, when a child uses a bicycle as a horse she does so to represent a horse, but she doesn't mistakenly take it to be a horse. Similarly, with linguistic representations, (1) may be used to express a possibility, something similar to (2) may only be used to express a representation of a possibility. To account for this distinction I find it useful to use the technical notion of a "prop-sentence" which, unlike a normal sentence, uses objects as props (or stand-ins) of other objects. As such, prop-sentences use objects as representations of objects. Thus, they convey representations of possibilities and not possibilities.

My claim is that even though nothing satisfies the sentence in (2) there is some or other CD that satisfies the sentence in (3): i.e., that which takes a relevant object as a prop for representing Holmes.

(3) $\text{Holmes}_{\text{prop}}$ is a detective.

The decoupler produces the interpretation(s) needed to satisfy a prop sentence such as 3. It is such interpreted representations that the cognitive theory proposes as the sort of content associated with empty referential terms. Speakers communicate contents such as these when asserting sentences involving empty names, and it is contents such as these that speakers may use to form their beliefs, desires, etc. and ultimately guide their behavior.[2]

I said (3) may be used to *represent* a possibility. This must be distinguished from what is *expressed* by (4).

(4) There is a possibility of a detective named "Holmes."

The possibility *expressed* by 4 is not the same as the possibility *represented* by (3). While (4) expresses the possibility *that some individual is both a detective and named* "*Holmes*," (3) *represents* the possibility *that Holmes is a detective*.

To understand this one must remember that prop sentences, such as (3), do not have semantically determined truth conditions. They are not *strictly speaking* true or false, but they do have a cognitively associated content and truth conditions determined by the relevant CD. If you want, we can distinguish between what (3) *expresses* and what (3) *represents*. It expresses the possibility *that there is an object that is used as a prop or stand-in for Holmes and it is a detective*. Yet, what does the explanatory job here is the associated CD that is a vehicle of thought that represents the possibility *that Holmes is a detective*.

It is also useful to distinguish this account from some recent proposals owed to Sainsbury (2005) and Perry (2001). According to Sainsbury (2005), a sentence like (2) is meaningful because all names (empty or not) have two kinds of semantic content: their referent and their referential condition. So, for example, "Aristotle" has the following two contents:

Aristotle himself; and
The condition that $\forall x\{(\text{“Aristotle”}\textsc{r}x) \equiv (x = \text{Aristotle})\}$ where "ʀ": reference relation.

An empty name, such as "Holmes," has only the latter kind of semantic content on Sainsbury's view. It should be clear that this account takes (2) to

convey something different from what (3) conveys, namely, something more like (5).

(5) For every object, it is the referent of "Holmes" if and only if it is identical to Holmes, and whatever satisfies this condition is also a detective.

Sainsbury's semantic account has it that all simple sentences involving empty names are *strictly speaking* false (this can be problematic, see section 6.3). According to the cognitive theory, I defend this is simply not the case. The main difference between Sainsbury's view and my own is that on the latter meaningful sentences involving empty names may have no semantic content (not even of the referential condition kind). What makes their use intelligible is the cognitively associated content (i.e., a CD).

Perry (2001) proposes that all referential terms have two kinds of content: referential and reflexive. The former is determined to be the referent of the term; the latter is the set of conventions associated with the use of the term that allow speakers to identify the referent. This distinction is quite clear in the case of demonstratives: they have a referent and a character, the latter constitutes its reflexive content.

Since empty names have no referent, they only get their reflexive content and, thus, the sentences involving empty names only get reflexive truth-conditions. Perry (2001) takes (6) to express the reflexive truth conditions of a sentence such as (2).

(2) Holmes is a detective.
(6) That the convention exploited by uses of "Holmes" in (2) has an origin and it is a detective.

It should be clear that the possibility *expressed* by (6) is not the same as that *represented* by (3) (see above). Furthermore, this account also seems to differ from my own in the same way that Sainsbury's: it is a semantic one. Like Sainsbury's (2005), Perry's view seems to imply that, like (6), all simple sentences using empty names (i.e., excluding negative existential assertions) are *strictly speaking* false.

Let me now describe the cognitive account of empty names by presenting its central assumptions and theses. First come the assumptions:

Semantics: in virtue of lexical semantics, all referential uses of names do *not* contribute descriptive information to the content of utterances in which they appear.

Pragmatics: all referential uses communicate a non-proffered piece of information: that there is a referent of the name (some take this to be a presupposition).

These assumptions are based upon independent research. The semantic assumption is based on empirical research from cognitive psychology and psycholinguistics according to which proper names do not have descriptive *semantic* content (see García-Ramírez and Shatz 2011). I refrain from appealing to non-descriptive semantic content beyond that of the referent (see Sainsbury 2005). I take it that the simplest account of the semantics for names takes them to have a referent and nothing more. Since the account I give is not semantic, the phenomena of empty names do not constitute reasons to add anything else to the semantics. This, of course, does not imply that there may not be other reasons that I have ignored.

The pragmatic assumption, in turn, seems to me to be uncontroversial. There is, to my mind, a lot of agreement on this (see Heim 1982; Roberts 1998 and 2004).

Now, these are the theses:

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

Flexibility: two attitudes, the content of which is cognitively determined (as above), may have the same associated representation (a CD) even if the latter have been (cognitively) assigned different props as referential conditions.

Depictions: the relevant contents associated to the use of empty names are determined by cognitive depictions (CDs) that result from the decoupler's assignment of a prop as the referent of the relevant term. The resulting CD satisfies a prop version of the relevant sentence.

I hope the previous discussion is enough to illustrate why I defend the theses, as I take them to be my contribution to the debate on empty names. As I described the relevant notion of "content" above, the lexical semantics for names plays no determinant role. Why is it, then, that an anti descriptivist (perhaps even referentialist) semantics for names is part of the assumptions? The answer is simple. The phenomena of empty names have long been taken to pose a problem for anti descriptivist theories of names (see Brock 2004): they must either accept descriptive information in the interpretation or offer a homogeneous view according to which empty and non-empty names either have different contents or determine the interpretation in different ways. As I said, there are good (independent) empirical reasons for not accepting descriptivist views of names. Thus, if the cognitive theory of empty names is right, empty names do not pose a serious problem for anti descriptivist theories of empty names—see Soames (2002); Salmon (1998); Braun (2005);

Reimer (2001b); Adams and Stecker (1994); Taylor (2000); Walton (2000); Sainsbury (2005); and Stalnaker (1978).

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 lexical semantics *alone*. So we need the cognitive part of the theory to explain how such uses *are* intelligible. I have offered a psychological account of the mechanism involved in uses of empty names. Even though it relies on current empirical research on pretense, this account is still preliminary. More needs to be said about the relation between pretense and the development of mental representational states. The central claim, however, remains. 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 "cognitive depictions" or CDs.

6.3 SOLVING PHILOSOPHICAL PUZZLES

The literature traditionally distinguishes among four problems associated with empty names (see Reimer 2001a, and Braun 2005): meaningfulness, truth-value, content, and attitude reports. As I will show in this section, the cognitive theory can solve all these problems and more.

Consider first the problem of meaningfulness. Speakers seem to think that the sentences they use are meaningful even though they use empty names -at least when they do not know that the name they use is empty, as in Le Verrier's case. The cognitive theory has a straightforward explanation. Since there is no salient object perceptually available at the time of use, the decoupler is triggered to provide an interpretation. This provides the subjects with an interpretation that makes the use of the name meaningful. They take the speech acts to be meaningful in virtue of cognitively interpreting them to be so. Speakers (and hearers) find referential uses of empty names to be meaningful because they imagine these names to have a referent. The same account explains how third parties may understand empty names without knowing them to be so. 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 other speakers are talking about, yet she finds the utterance meaningful. Unless there is a salient object perceptually available for the third party to associate with "Hamlet," the decoupler will be triggered and the name will be given an interpretation, most probably by means of a prop.

Speakers also seem to think that assertions using empty names do have some or other truth-value. This is the truth-evaluability problem. Le Verrier's assertion of (7), for example, was either true or false. Furthermore, we all

seem to think that (7) is false. The cognitive theory explains why we intuit that the sentence is truth-evaluable. The sentence gets an interpretation by the decoupler, one that may be said to satisfy the prop sentence in (8). The prop sentence is, in turn, truth-evaluable.

(7) Vulcan perturbs the orbit of Mercury.
(8) $\text{Vulcan}_{\text{prop}}$ perturbs the orbit of Mercury.

Since, as a matter of fact Vulcan does not perturb the orbit of Mercury, the possibility *represented* by (8)—*that Vulcan perturbs the orbit of Mercury*—is not an actual one, and so the sentence may be said to represent something false.

Attitude ascriptions involving empty names in the relative clause are also problematic. According to the theory here proposed it is CDs, *qua* vehicles of thought, which direct the attitudes associated with empty names. They are fine-grained enough to be psychologically explanatory 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 (i.e., their props), would express singular propositions. But how exactly does this work for belief reports? Here are some examples:

(9) Sally believes that Santa is coming tonight.
(10) Sally believes that Holmes is coming tonight.
(11) Le Verrier believed that Vulcan is a planet.

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 does not take speakers to be pretending (or to ascribe pretense states) when using empty names. As the examples show, CDs allow us to distinguish between beliefs about Holmes and those about Santa, even though "Santa" and "Holmes" are both empty. More specifically the embedded sentences in (9), (10), and (11) are associated with the interpretations that satisfy the prop sentences (12), (13), and (14), respectively.

(12) $\text{Santa}_{\text{prop}}$ is coming tonight.
(13) $\text{Holmes}_{\text{prop}}$ is coming tonight.
(14) $\text{Vulcan}_{\text{prop}}$ is a planet.

There are, of course, serious—non-pretense or fiction involving—uses of empty names. These would seem to pose a threat to any theory of empty names based on the notion of pretense. Yet, this is not the case of the theory I am proposing. Consider Le Verrier's famous case. He postulated the

existence of a planet in order to account for 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 by its lexical semantics. At the same time, Le Verrier postulated the existence of a planet in order to explain the disturbances in the orbit of Uranus. He called this hypothetical planet "Neptune." It turned out that there was such a planet, so "Neptune causes disturbances in the orbit of Uranus" turned out to have a truth-value determined by its lexical semantics. Le Verrier had exactly the same kind of evidence for postulating the existence of Vulcan as he had for Neptune, yet the difference between one and the other could not have been wider. A proper account of empty names should be able to explain this without postulating any difference relevant to Le Verrier.

If, on the one hand, Le Verrier did have evidence offering a perceptually available salient object that he associated with the relevant names or he didn't. Either way, it was the same for both names. Consider the first option, Le Verrier did associate a perceptually available salient object with "Vulcan" and (let's suppose) a different one with "Neptune." If this were the case then Le Verrier did interpret both names without requiring the use of the decoupler. The difference lies in the fact that for the case of "Neptune" he was correct in associating such perceptually available object with the name, but he was mistaken with "Vulcan." The perceptually available salient object turned out not to be a planet. Thus, the difference is owed to a fact that Le Verrier ignored.

If, on the other hand, Le Verrier had no such perceptually available salient object to associate with the relevant name, then the decoupler is triggered (see section 6.2). Thus, "Vulcan" and "Neptune" are not at any degree different for Le Verrier. The difference between both names lies in the fact that "Neptune" turned out to have a referent and "Vulcan" did not, but this is something Le Verrier ignored.

Le Verrier did believe the representation he imagined, 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 inadequate, incorrect or unfit.

Let me now address cases of errors and hallucinations. 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 lexically 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 is the referential condition of "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 head and can be used as a prop by the decoupler. 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 an incorrect or unfit belief (see Byrne 2009).

Friend (2009) has challenged theories of empty names, and particularly descriptivist ones, by demanding the satisfaction of the requirement of singularity: speakers behave as if there is a unique thing that they are talking and thinking about. The evidence comes from *coidentifications* and *counterfictional imaginings*. Coidentification is "the phenomenon of thinking and talking about the same thing, even when there is no such thing." Counterfictional imagining happens when subjects imagine what would happen to a given character had the fiction been different.

Both phenomena suggest that there is something that subjects have in mind and that we must identify this content independently of the properties subjects associate with the relevant characters. If, for example, it is part of what Holmes is that he is a great detective, it seems hard to explain how two subjects may be thinking about Holmes even if one does not take him to be a great detective; or how one may engage in what would have happened to Holmes had he not been a detective.

The account of the content associated with empty names that I have given is able to meet both challenges. The CD associated with an empty name does not depend on the properties associated with the use of that name. Holmes, for example can be identified as what "$\text{Holmes}_{\text{prop}}$" represents. Thus, two subjects may be said to be thinking about the same thing even if they do not associate the same properties with it. The same goes for counterfictional imaginings.

Now, it has been objected against pretense (see Walton 2000) and other (Sainsbury 2005) accounts of empty names that they cannot explain uses of fictional names outside of fictional contexts. To illustrate, consider (15)-(18) (owed to Sainsbury 2010):

(15) The Greeks worshipped Zeus.
(16) Holmes is famous.
(17) Anna Karenina is smarter than Emma Bovary.
(18) Santa is a fictional character.

Most accounts have a problem with these sentences because there seems to be no fiction operator or fictional pretense involved. So we cannot explain the truth of these sentences in terms of such contextual features. Unless they postulate fictional characters, most theories lack a representation that may be said to be true and that may be properly associated with a subject's

non-fictional attitudes and assertions. The cognitive theory has some such representation: a CD.

Perhaps the most long-standing, classical problem associated with the use of empty names is the problem of intentional intexistence. 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, 88–89)

This paragraph 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 2014). The first thesis defines mental states as intentional: 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.

The cognitive theory explains how speakers can be said to have thoughts about Santa, Hamlet, Vulcan, etc., by having thoughts about something distinct from their mental states themselves, i.e., a CD. Thus, Brentano's second thesis is accounted for: by having mental states directed to CDs, a speaker may be said to have thoughts about something that does not exist: the possibility represented by the relevant CD. There is no need to look for special entities that do not exist. All we need is a relevant representation the interpretation of which does not require anything but existing objects. That is exactly what a CD is. These objects exist; they are vehicles of thought required for humans to pretend (see Leslie 1987 and Nichols and Stich 2002). Thus, there is no need for inexistent objects.

Furthermore, CDs are distinct from the mental states—beliefs, desires, etc.—that are directed to them. Thus, Brentano's first thesis is also accounted for. 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 from the mental state itself.

I think this account of intentional inexistence does better than its rivals (see Kriegel 2007). But I don't have the space to defend it here. For some independent evidence on behalf of this view, see Gómez (2008), who offers an evolutionary account of pretense by associating it with the appearance intentional inexistence.

Serious uses of fictional empty names such as "Hamlet" naturally give place to ontological queries. Consider the Shakespeare scholar whose life is dedicated to the study of Hamlet. What is she studying? Fictional entities? Some, like Salmon (1998), are happy to say yes. Others, like Thomasson (2003), take them to be cultural artifacts. The theory I have proposed suggests, however, a negative answer: 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.

Unless they are about actual things (as they may well be), all we need to account for literary works is a set of cognitive depictions that may all be part of a common causal history. If scholar A's use of "Hamlet" is causally related to scholar B's use in the relevant way—they are both talking or thinking 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 (Everett 2000).

Walton (1990), among others, points out one further problem. Fictional discourse and fictional truth, he argues, are importantly distinct from non-fictional discourse and truth. Part of the difference has to do with the fact that, as Walton puts it, fictional claims are self-warranting. It is because Conan Doyle says so that Holmes is a great detective, but it is not because Darwin says so that species evolve. Determining exactly how it is that fictional discourse works is a complex and vexing issue. Walton (1990)—and Friend (2009) following Walton (1990)—claims that it does so by prescribing imaginings.

If this is so, then so be it. The cognitive theory can fit this account pretty well: fictional discourse prescribes imaginings and the subject complies with this prescription by means of the decoupler. One need only add that the decoupler explains how the prescribed imaginings are realized. Since the decoupler is an independent mechanism of its own, the account can be suitably extended to account for non-linguistic fictional representations in general.

To close this problem-solving section, consider the well-known problem of negative existential assertions. It seems that empty names may be used to assert literal truths outside of fictional contexts when used in negative existential environments such as "X does not exist." The tradition has offered three distinct ways of accounting for them: as metalinguistic claims (see Walton 2000; Thomasson 2003; and Stalnaker 1978), as quantified phrases (see Russell 1905), and as literal truths (see Soames 2002; and Sainsbury 2005). The cognitive theory is compatible with all three accounts.

On the metalinguistic view, negative existential assertions convey information about the referential term used in the negative existential claim. The referential term used is not strictly speaking empty: it is used to refer to itself. If so, there is nothing for the cognitive theory to explain. It is completely consistent with such metalinguistic account.

On the quantificational view, negative existential assertions are interpreted as quantified phrases. There is nothing like an empty referential expression involved. Thus, there is nothing for the cognitive theory to explain, so it is consistent with a quantificational theory of negative existential assertions (but see Walton 2003).

On the literalist view, negative existential assertions literally predicate lack of existence of an object. If so, one may take a suitable CD to be the content of the assertion "X doesn't exist." If "exists" requires a special interpretation, this may be accounted for by the decoupler's process of interpretation (see section 6.1).

6.4 CLOSING REMARKS

I like to see the cognitive theory of empty names as having a mix of strength and parsimony. Its strength relies on the fact that it explains all the relevant phenomena (see section 6.3) while allowing for a homogenous interpretation for referential uses of names. Its parsimony is owed to its not postulating special ontology (e.g., inexistent objects, fictional entities) or implausible psychological mechanisms. It appeals 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. In this sense, the theory is well supported. It is not only consistent with the empirical research on pretense and fictional reasoning, it stems from such research.

The theory does claim that empty names, in particular, demand more than just linguistic resources. But this, of course, is a feature that we should expect to find in any theory resulting from the endorsement of *open compositionality* and its *cognition-first* methodology. As such the theory of empty names here presented gains support from the arguments and evidence for *open compositionality* presented throughout this book. I believe these are good reasons to take the cognitive theory of empty names as a serious alternative.

In the following chapter I will consider one last application of the *cognition-first* methodology, this time with the aim of shedding some light over the nature of moral discourse and practice. Together, I hope these theoretical products of *open compositionality* are enough to show how this

general view of natural language works, how it can be applied to account for distinct phenomena, and why the *cognition-first* methodology is a powerful one even though it is not as simple, transparent, and easily applicable as the compositionally closed methodology.

NOTES

1. This chapter was originally published as García-Ramírez, E. 2011. "A cognitive theory of empty names." *Review of Philosophy and Psychology* 2, 785–807. Thanks are owed to the editors at Springer for granting the copyright permission.

2. The claim that the same mechanism may inform belief-like and pretense-like mental states may seem strange to the reader. They seem to be radically different kinds of mental states. The former is meant to produce accurate representations of the environment while the latter may produce deviant ones. But there are important similarities among them. First of all, they are both meant to represent the environment. Even if deviant, subjects "play along" with what they pretend. Second, they may both be part of inferential processes, pretense states may be part of the development of a game of make-believe. Third, 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).

Chapter 7

Moral Discourse, Moral Cognition, and the Language Analogy

Is there such thing as moral knowledge? If so, how is it possible? Philosophers have addressed these questions at least since Plato. Contemporary philosophy has placed them at the center of what is known as metaethics or, more specifically, the metaphysics, epistemology, philosophy of mind, and philosophy of language of moral discourse, thought, and practice (see Miller 2014; Jacobs 2002; Horgan and Timmons 2006; Darwall, Gibbard, and Railton 1997). Intuitively, the first question has a positive answer. There is moral knowledge and humans have plenty of it. We know, for example, that causing unnecessary harm is bad, that helping others is good, that suffering should be avoided, and so on. It is the second question that has proven to be highly problematic, so much that is has cast doubt over the correctness of the intuitive positive answer to the first question. If there is moral knowledge, as the intuitive answer claims, then the so-called "moral truths" must correspond with some or other moral fact or moral property of which the said moral truth is a truth. This, however, generates problems of all kinds.

First, the metaphysics involved seems dubious. There seem to be no normative or moral facts in the world, there are just facts (see Railton 1986; Brandt 1979; Brink 1989). So what kind of facts and properties are we talking about when we talk about moral knowledge? Second, there seems to be no perceptual access to moral facts. Moral facts and properties would seem to demand a special methodology (see Sturgeon 1985; Nelson 1996; Campbell 2014). What sort of epistemological access do moral facts demand? Third, if our epistemological access is not a simply causal / perceptual one, then how is it that moral terms such as "good" and "bad" end up referring to them? A simple, causal semantics for moral terms would seem to be of little use (see Schroeder 2012; Wedgwood 2007). What do moral terms denote and how do they get to do so? Fourth, intuitively, moral judgments have a motivational

force (see Hume 2007; Dancy 1993; Smith 1994). Sincerely judging, for example, that helping others in need is good will naturally lead us to actually help others in need whenever possible. But if moral knowledge is a matter of knowing the facts and properties of the world how can it motivate us? What kind of mental states are associated to moral knowledge so that it ends up having such motivational force (see Campbell 2007; see also Kumar 2015)? Or is there something about moral facts and properties that they sufficiently motivate human behavior (see Mackie 1977)? Fifth, and last, if there are moral facts and properties, the distribution of moral judgments and moral views should be somewhat similar across cultures. Yet, the evidence shows that there is more cultural variation and group-relativity than uniformity in the distribution of moral judgments (see Super and Harkness 1997 and 2002; Whitting and Edwards 1988; Knight and Carlo 2012; Brenick and Killen 2014).

Naturalist moral realism stands out among the views that intend to solve these problems. It vindicates the intuitive claim that there is moral knowledge while claiming that moral facts and properties are nothing over and above the moral facts and properties of empirical science. I believe moral naturalism is our best alternative for defending moral knowledge, yet it faces some serious problems, especially with respect to the semantics of moral discourse. If moral properties are natural properties, then moral terms should refer to the latter. However, if such were the case, it is argued, moral terms should be substitutable for the corresponding terms for natural properties. This, however, does not seem to be the case. As Moore (1903) famously points out, it is perfectly acceptable to ask "I know that doing such and such is F, but is it good?" where "F" stands for the corresponding natural property allegedly denoted by moral terms. However, if the naturalist semantics is right, such questions should be unacceptable as they are, according to the naturalist proposal, semantically equivalent to asking "I know that doing such and such is good, but is it good?" (see Railton 1986, 204–207)

The above problem is commonly known as the "open question" problem. By now it should be clear to the reader that this problem is intimately related to the closed view of language and the assumption of *strong compositionality*. It makes sense to worry about the meaning of moral terms by looking into how the meaning of these and other co-designating terms behave with respect to the meaning of questions only if we assume that the content of the latter (i.e., a complex expression) is determined by the meaning of the expression parts (see chapter 2). Once we give up the closed view, together with *strong compositionality*, and endorse open compositionality and the cognition first methodology, we gain access to a rich and varied set of resources beyond syntax and semantics that may help us solve the open question problem—and potentially many others—threatening the success of naturalist moral realism.

According to open compositionality, natural language is a supermodular cognitive ability useful both as a communicative tool and as a platform for higher order thought. As I will show in this chapter, language plays both roles with respect to moral cognition. It helps communicate moral judgments and engage in moral inquiry, while facilitating moral understanding. Moral knowledge, as I will argue, is better understood by analogy with human knowledge of language. This language analogy, I believe, is a very fruitful one, as it naturally delivers the basic elements of an account of moral knowledge based on a naturalist ontology of moral facts, an ordinary epistemology for moral knowledge, and a simple semantics for moral discourse. The resulting view also suggests an internalist view of moral motivation, an account of why there are robustly objective moral truths, and an explanation of why there is so much cross-cultural moral variation.

According to the cognition first methodology (see section 4.4) if we want to understand how moral discourse works, we must first understand moral cognition. To do so I will look into studies of moral cognition and moral cognitive development. To have a clear view of how understanding moral cognition may help us understand moral knowledge, I will begin in section 7.1 by describing the core elements of naturalist moral realism. In section 7.2, I will present a brief yet detailed account of moral cognition according to our best empirical studies on moral cognitive development. Based on this evidence, in section 7.3, I will develop what I call the "language analogy," the claim that moral and linguistic cognition are substantially similar to each other and should be theoretically approached in parallel ways. In section 7.4, I show how we can derive substantial notions of moral objectivity, moral conventionality, and moral cultural variation from this account. In section 7.5, I present what may be considered the blueprint of a theory of moral knowledge that avoids traditional philosophical problems by understanding moral knowledge as associated to moral competence—i.e., acquiring, sustaining, and developing moral cognitive abilities—and present a brief account of what this moral competence consists of. I conclude by showing how the suggested account avoids the most common objections against naturalist moral realism coming from metaphysics, moral motivation, and objectivity.

7.1 THE CHALLENGES OF NATURALIST MORAL REALISM

Although, as Railton (1986) mentions, there are many different theories that claim to be versions of naturalist moral realism, for present purposes I will assume that the view is best characterized as defending the following three claims. Naturalist moral realism claims first, that there is in fact

moral knowledge; second, that the moral facts and properties warranting such knowledge are nothing over and above the natural facts and properties of empirical science; and, third that such moral facts and properties are real facts and properties in the sense of having a mind-independent reality. The challenge for a naturalist moral realist is to find "a plausible synthesis of the empirical and the normative." (Railton 1986, 163) This synthesis refers to an acceptable theoretical equilibrium whereas both, the normative and the empirical, aspects of morality are well accounted for. A naturalist moral realist's goal is, thus, to show how our best empirical theories allow for normative functions to take place in our world. To achieve this difficult theoretical task is, thus, to offer an account of moral truth and its warrant that is compatible with a naturalist moral epistemology—i.e., an account of how these truths are known that does not postulate special or *sui generis* epistemic access to such truths. As Railton (1986) puts it, moral truth and its warrant must be such that there is some "reliable causal mechanism for moral learning." (Railton 1986, 171)

Railton (1986) sets a high standard for any version of naturalist moral realism to constitute a satisfactory account. First, the facts and properties postulated by the theory must exhibit two important signs of reality and behavioral guidance, dubbed "independence" and "feedback":

> *Independence*: it exists and has certain determinate features independent of whether we think it exists or has those features, independent, even, of whether we have good reasons to think this;
>
> *Feedback*: it is such—and we are such—that we are able to interact with it, and this interaction exerts the relevant sort of shaping influence or control upon our perceptions, thought, and action. (Railton 1986, 172)

Second, the proposed account must also show how, in virtue of the postulation of such and such moral facts and properties, the theory can perform a much-needed explanatory function, accounting for our everyday experience. Furthermore, if the proposed theory is to be seriously considered as a satisfactory one, it must show how this explanation is compatible with available empirical evidence.

Now, while meeting these standards any naturalist moral realism must somehow distinguish moral from non-moral facts, since only the former may warrant the truth of moral judgments while also having practical or guiding force upon us. Railton is skeptical that this can be done by means of a psychological strategy that postulates a special form of moral reasoning and so, he argues, we must "shift to ontological ground." Instead, Railton (1986) offers a metaphysical account of the nature of moral facts within naturalist

moral realism (see also Brandt 1979; Copp 1995; Brink 1989; and Sturgeon 1985).

Railton (1986) understands moral rightness as "what is rational from a social point of view with regards to the realization of intrinsic non-moral goodness." (Railton 1986, 191) Social rationality is idealized in this notion, as is non-moral goodness understood as the objective interest of an agent, what a subject "would want himself to seek if he knew what he was doing." (Railton 1986, 177) At the end of the day, moral rightness supervenes upon an agent's desires and other attitudes, her circumstances, and the dispositional properties of the objects of interest to the agent in the circumstance. At bottom line, moral rightness is a relational property that holds between the agent, the circumstances, and the relevant objects. Understood as agential interests approved by ideal social rationality, the postulated moral property / fact satisfies the requirements. It has existential independence; agents do have interests independently of whether we think they do. And there is feedback, as agents may interact with such interests by finding out through experience which ones among them are members of the relevant objective subset, and acquiring new interests accordingly while modifying their behavior. Railton (1986) argues that this feedback also accounts for its explanatory function. Moral rightness is necessary to explain how an agent's modified behavior and changing interests constitutes an improvement of her rationality (see Railton 1986, 188–89).

The trouble for Railton's (1986) account comes from the proposed semantics for moral terms. Railton (1986) defines the term "moral rightness" as denoting a set of intrinsic objective interests approved of by idealized social rationality. If we pair this with an implicit assumption of the closed view's *strong compositionality*, it follows that "moral rightness" contributes something like "being an objective interest that would be approved of by idealized social rationality" to the meaning of any complex expression. This is, intuitively, mistaken. When ordinary competent speakers utter something like "Helping others is morally right." They do not seem to be saying something like "Helping others is an intrinsic objective interest that would be approved of by idealized social rationality." Railton (1986) admits this is problematic, as it makes the theory vulnerable to Moore's "open question" arguments (see Railton 1986, 204–7). One can know that helping others is an intrinsic objective interest that would be approved of by idealized social rationality and still ask if it is the morally right thing to do.

More trouble comes from the motivational function that moral properties are supposed to have. Unlike natural properties, moral properties are a source of motivation in the sense that by merely judging that seeking *o* is morally right an agent acquires, *ipso facto*, a motivation for seeking *o*. The source of such

motivation is not to be found in the reduction base of moral properties, for these are constituted only by ordinary, non-normative natural properties. The only potential sources of moral motivation left are the idealized notions of an objective interest and of an ideal social rationality. Railton thinks that the very idea of objective interests is somehow a source of motivation. If one is to learn that one would still want *o* if one were to be "fully informed and perfectly rational," then presumably this would motivate one to want *o*. But this account seems to presuppose, rather than account for, moral motivation. One will be motivated to have objective interests, and to make one's goals such that they would be approved of by ideal social rationality, only if one is *already* interested in doing what is the morally right thing to do. It is at least not obvious that just the thought of a fully informed and fully rational agent wanting *o* will make *me* want *o*. And it is even more doubtful that I need to reflect on these matters in order to be motivated to do what I consider to be the morally right thing to do.

Thus, there is room for a more satisfactory synthesis between the empirical and the normative aspects of morality in the search for a naturalist moral realism. I believe some such account is forthcoming once we adopt the *cognition-first* methodology. As I will argue, a proper understanding of moral cognition delivers an account of both moral epistemology and metaphysics that avoids the semantic and moral motivation problems facing Railton's theory. This methodological shift requires a move from the ontological ground of objective interests into the psychological ground of human cognition.

7.2 MORAL COGNITION AND DEVELOPMENT

Since Piaget (1932) and Kohlberg (1969) the study of human moral cognition and development has vastly advanced, following in part the extraordinary progress made by what are nowadays known as the cognitive sciences. Human moral cognition is now the focus of study within a rich interdisciplinary setting that includes cognitive psychology, evolutionary biology, anthropology, and neuroscience (among others). From PET scans and fMRI tests to behavioral studies—with humans and non-humans—moral cognition is under close empirical scrutiny. There is, as one would expect from such a varied interdisciplinary approach, an enormous amount of research on the topic (see Killen and Smetana 2006; Brugman, Keller, and Sokol 2013; Nucci, Narvaez, and Krettenauer 2014; and Lapsley and Carlo 2014). It is not my goal to present an all-encompassing review of this thriving field, not even a brief one. Instead, I will simply focus on a few outstanding features of moral cognition and development that will prove to be of great significance for the purposes of this chapter.

Evolution of Morality

Human morality has an evolutionary history. Studies with primates have shown that chimpanzee's develop and enforce prescriptive norms for certain socially salient behavior, such as mating and caring for infants. Whenever such a rule is violated, third parties usually give their attention to the victims and identify the transgressors. Together, victims and third parties, commonly punish the transgressor (see de Waal 1991 and 1997). Aggressive, reconciliatory, and altruistic behavior have also been observed in non-human primates (see de Waal 1998; Hauser 2006). This suggests that moral cognition is a unique and salient element of human cognitive endowment just as much as the human visual system. It also suggests that moral knowledge, whatever it may turn out to be, is at least partly the result of natural selection and not purely a matter of experience-based scientific-like knowledge.

Based on these evolutionary studies, as well as on studies from multiple other sources—i.e., social, cultural, and biological—Haidt forcefully argues that moral reasoning is a post-hoc construction that usually follows once a moral judgment has previously been made. Moral reasoning, whenever it takes place, appears to be working as a defense strategy to support the already formed moral judgment. Moral judgments are automatic, not the result of much reflection (see Haidt 2001 and 2007). Even so, moral judgments play a central evolutionary and cultural role as "social binders." Aside from the automaticity of moral judgment, the evidence also shows that emotions are more significant than reasons when it comes to determining moral judgment and consequent behavior (see Haidt 2001).

Assuming that there is moral knowledge, it must be of the sort that can be—at least partly, if not substantially—evolutionarily endowed. It must be capable of being exercised in an automatic, fast, and simple fashion in a way that resembles the use of fast and frugal heuristic principles in language processing (see section 4.3). Moral knowledge also appears to demand emotional intelligence, more so than scientific knowledge. Some have argued that these features are enough to foreclose the chances of any such thing as moral knowledge, for knowledge cannot result from such an emotional, intuitive, and frugal cognitive process as described (see Kahane 2011 for a review of such evolution-based objections to moral knowledge). Others, such as Railton (2014), defend the possibility of moral knowledge by arguing that, when properly understood, intuitions and the human affective system are not incompatible with knowledge, even scientific knowledge. Like Railton (2014), I believe the evidence does not show that there is no moral knowledge. What the evidence shows, I think, is that moral knowledge is of

a special kind, one that is substantially determined by the nature and state of human cognitive endowment.

Prelinguistic Moral Principles

Carey (1997) offers a useful criterion for identifying universal cognitive principles needed for a given human cognitive function to perform and for distinguishing them from representations that do play a role in the said cognitive function but are in fact cultural constructions that differ from population to population. To do so, Carey (1997) argues, we must first identify the candidates for universally cognized representations and then "establish whether these articulate the mental representations of prelinguistic human infants." (Carey 1997, 37) Insofar as the identified representations are prelinguistic, there is good reason to think that they are universal and perhaps also innate. With respect to those representations that are found to play a role in the same cognitive functioning but do not express infant understanding, we must determine "when these conceptual resources become available to children, and explore the mechanisms by which they do so." (Carey 1997, 37)

I believe Carey's (1997) criterion may be of great use when it comes to understanding the nature of human moral cognition and how it develops from infancy into adulthood. Multiple studies have identified what, following Carey (1997), may be understood as universally cognized moral representations, that is, moral concepts or principles that are common to all human individuals across populations (see Cushman, Young, and Hauser 2006; Hauser 2006). I will call these "prelinguistic moral principles," which include principles such as *suffering is bad*; *helping others is good*; *hindering others is bad*; *each one must receive what is deserved*; and *it is good to be empathic*. Whether some or other moral principles are part of this universal, prelinguistic set of moral representations is, of course, an empirical matter. So far, these are some of the principles for which there is independent empirical evidence.

Martin and Clark (1982) replicated studies on newborn sensitivity to suffering. The study involved 70 newborns, which were presented with tape-recorded crying of other infants. Calm infants cried in response to the crying of other infants, but showed no response to their own crying, and ignored the crying of a chimpanzee and that of an older child. Crying infants, on the other hand, continued to cry when presented with the crying of another infant, but almost stopped crying when presented with a recording of their own cry. The evidence shows human newborns are equipped with sensitivity to expressions of human suffering and with empathy for others.

In a study with six- to ten-month-old infants Hamlin, Wynn, and Bloom (2007) found that from early on human infants engage in social evaluations of unrelated third parties. Infants were presented with an animate character (the

climber) that was either helped or hindered by other characters in its attempt to climb a hill. The experiment used a choice paradigm for determining preference—infants reach out at what they prefer—and a violation of expectation paradigm for determining surprise—infants look longer at surprising events. Infants robustly reached for the helper character over the hinderer, and were surprised to see the climber approaching the hinderer afterwards. A further experiment included a neutral character that simply did not interact with the climber. Infants reached more for the helper than for the neutral character, yet preferred the neutral character to the hinderer. "That is, infants were both drawn towards helpers and independently inclined to avoid hinderers revealing positive and negative evaluations." (Hamlin, Wynn, and Bloom 2007, 558) Further studies (see Hamlin, Wynn, and Bloom 2010) show that this capacity for social evaluation, although with a privilege for negative social information, is already present at three months of age. These studies suggest that human infants are equipped with a capacity for assessing the actions of others, as well as with a prelinguistic principle that positively values helping over hindering. Hamlin and colleagues argue that this capacity for social evaluation, given how early it appears in infancy, can be seen as a biological adaptation that "may also serve as a foundation of a developing system of moral cognition." (Ibidem)

This understanding of the value of *collaboration* seems to be substantially refined by the first year of age. In a study with 14-month-old infants Henderson and Woodward (2011) found that infants have a complex understanding of collaborative work. They seem to know that collaboration is complementary yet critical to achieving the goal, and that there is no collaboration unless there is a causal relation between the actions of the helper and those of the agent in need of help.

Between the second and third year of age human infants already understand *retribution*. When resources are scarce they tend to distribute them unequally between helpers and hinderers, with a preference for the former. Yet preschoolers also exhibit a preference for equal distribution of resources when these are plentiful (see Kenward and Dahl 2011). A separate study by Hamlin, Wynn, Bloom, and Mahajan (2011) shows that this understanding is rooted in an early ability to assess moral actions and their agents not only by considering the moral worth of both the agent and the recipient of the action. Subjects were presented with a scenario including two characters; one intending to perform a difficult action while the other character either helps or hinders the first one. After observing these interactions, subjects were presented with a second scenario including as agents the helper and the hinderer of the previous scenario, both of which were performing an action for which they needed help and they were both either helped or hindered. The results show that by eight months of age human infants prefer those who help prosocial

agents—i.e., those that were helpers initially—as well as those who hinder antisocial agents—i.e., those who were hinderers initially. The evidence shows that very young infants have a rather complex capacity for social and moral evaluation, showing sensitivity both to the agent's moral worth as well as to the context in which the action takes place.

As with collaboration and the value of helping versus hindering, *fairness* is another moral notion for which there appears to be prelinguistic preparedness. In a study with 12- to 18-month-olds, Geraci and Surian (2011) presented subjects with two alternative animations involving four characters, namely, a distributor of resources, two recipients, and a bystander. On one first scenario, the distributor would effect an equal distribution of goods in front of the bystander, while an unequal distribution would take place in the alternative scenario. Infants were then asked to select between the equal and the unequal distributors, and were presented with a final scenario in which the bystander would approach either one of the two distributors. Infants were asked to perform a manual task, reaching for their preferred distributor, and a looking task, observing the bystander approach either one of the distributors. Infants preferred the equal distributor to the unequal one in the manual tasks, and preferred looking at the bystander approaching the equal distributor than when it approached the unequal one. These results show that at 16 months of age human infants are sensitive to the outcome of distributive actions and take it into account when evaluating agents. The results also show that young infants already prefer equal distributors to unequal ones, and reason about others' preferences—i.e., the bystander—by assuming they will similarly prefer the equal distributor. Other studies (see Kuhlmeier, Wynn, and Bloom 2003) have shown that 12-month-olds are already capable of understanding goal directed actions and interpret an agent's behavior on the basis of her previous actions, even if they took place in physically distinct contexts. Kuhlmeier and colleagues (2003) argue that this ability is better understood in mentalist terms, whereby infants posit intentional or goal-directed mental states mediating between agents and their actions. Together these studies suggest that human infants are equipped with a prelinguistic principle of fairness (see Schmidt and Sommerville 2011 for further evidence).

These briefly reviewed studies show that human infants are equipped with a complex set of prelinguistic cognitive abilities that belong to the moral domain of human cognition. There is evidence of young infants possessing prelinguistic moral principles (see also Cushman, Young, and Hauser 2006; Pellizoni, Siegal, and Surian 2010), as well as a complex and context sensitive capacity for goal directed understanding of action and social evaluation. Following Hamlin, Wynn and Bloom (2010), it seems justified to claim that there are substantial, principle-based and evolutionarily endowed *foundations*

for human moral cognition (see also Wynn 2008). Given that these principles appear to articulate the mental representations of very young, prelinguistic human infants, there is good reason to think—following Carey's (1997) criterion—that they are universal. However many and distinct such prelinguistic moral principles turn out to be, all human beings are endowed with all and the same ones.

Together with the evolutionary evidence considered previously, the evidence of prelinguistic, foundational moral principles further substantiates the claim that human cognitive endowment determines and scaffolds moral knowledge. Moral competence is essentially constituted by an evolutionarily endowed cognitive foundation comprised, among other things, by evaluative abilities and principles of moral cognition (see Hamlin 2013). As I will show in what follows, this "innate moral core," as Hamlin (2013) calls it, also includes a rather complex mentalist understanding of agents and their goals.

Intentions, Emotions, and the Mental

Several studies on moral development have underscored the central role that the Theory of Mind (ToM) plays within competent moral cognition and vice versa. ToM has been postulated to account for the human ability to understand others, mainly their conspecifics, by attributing mental states (mostly representational ones) to them. ToM is considered to be a higher order cognitive ability, closely related with an early understanding of intentionality and necessary for humans (both infants and adults) to competently engage in social interaction and, importantly for our purposes, to competently acquire and develop natural language (see also Shatz 1994 and 2007b).

The initial workings of ToM appear early in infancy and seem to be effortless and automatic. In a study with seven-month-olds and adults, Kovacs, Teglas, and Endress (2010) tested subjects in order to determine if and how the beliefs of others were represented and how this affected the subject's behavior. Infants and adults were shown movies presenting an agent and an object (e.g., a ball). In some of the movies the object was placed in a certain location while the agent was present. In other movies the object was moved to a different location while the agent was absent. Adults were asked to signal as soon as they detected the object, seven-month-olds looking times were measured as a reaction to the location of the object. The results show that both, very young infants and adults immediately form a representation of the agent's beliefs, and that this representation was automatically taken into account when performing the relevant task. Adult subjects' reaction times were larger when the agent's beliefs differed from their own, and infant looking times varied similarly.

Other studies have shown that ToM is involved in human social evaluations since early on in infancy. A mentalist understanding of an agent as a helper / hinderer requires second order mental representations, as the goals of the helper / hinderer depend on the goals of a separate agent who may be helped / hindered in reaching her goals. For example, a helper is represented as having the goal of helping achieve someone else's goal *x*, *y* or *z*. Hamlin, Ullman, Tenenbaum, Goodman, and Baker (2013) tested whether ten-month-old infants were capable of having such a mentalist understanding as it is expressed in their social evaluations of agents. Subjects were presented with a puppet show including three characters, an agent *A* with the goal of grasping a preferred object *o* among a set of objects, and two other agents that play the role of a helper *H+* and a hinderer *H-* respectively. After the puppet show takes place, infants are asked to select between *H+* and *H-*, which are identical puppets except for a different clothing color. If *A* expressed its preference for *o* in the presence of *H+* and *H-*, infants would consistently prefer *H+* over *H-*. In contrast to this result, infants chose randomly between *H+* and *H-* whenever the latter where ignorant of *A*'s preference—either because *A* did not express its preference for *o* or because it did in the absence of *H+* and *H-*. These results strongly suggest that ten-month-olds are already capable of mentalist social evaluations, including a second-order understanding of agents as having goals that depend on those of others.

Now, aside from mental representational states, other mentalist forms of cognition and understanding appear to be also at play in moral cognition. Emotion understanding has been found to be a part of moral competence from early on in infancy. As I mentioned before, empathy has been found to be a component of human cognition from the very beginning, as evidenced by newborn empathic cry in response to the crying of other babies (see Martin and Clark 1982). This empathy appears to develop into more complex forms of affective response, with eight- to ten-month-old infants showing some affective and cognitive concern for others (see Roth-Hanania, Davidov, and Zahn-Waxler 2011), and 12 to 16-month-olds showing comprehension of the emotional state of others (see Davidov, Zahn-Waxler, Roth-Hanania, and Knafo 2013). Empathy and affective understanding becomes more complex as it interacts with a subject's cognitive and social development. In a study with four- to seven-year-olds Lagattuta (2005) found that four- to five-year-olds understood an agent's emotions in relation with her goals, whereas seven-year-olds and adults understood an agent's emotions also in relation to the rules and consequences that may have an influence on them (for a detailed review of prosocial behavior and development see Eisenberg, Spinrad, and Knafo-Noam 2015; and Carlo 2006). The evidence suggests, thus, that competent moral cognition involves both mental and emotion understanding. Studies of neuroimaging data seem to support a complex view of

moral cognition as involving not only complex mental representational and conceptual reasoning, but also a serious emotional involvement and cognitive control required for moral judgments (see Greene 2005a). For a general overview of the neuroscience of moral cognition, see Greene (2005b).

Other studies have found that theory of mind and emotion understanding closely interact as constitutive elements of social evaluation from early on in childhood. In a study of longitudinal data with children from their 3.5 to their 5.5 year of age Lane, Wellman, Olson, LaBounty, and Kerr (2010) found that both, the level of competence with ToM and with emotion understanding at age 3.5 predicted a more or less sophisticated moral reasoning at age 5.5, with high levels of ToM associated with greater reasoning involving the psychological needs of an agent, and high levels of emotion understanding associated with reasoning about the physical and material needs of agents. In general, the study found that the state of a child's competence with ToM and emotion understanding at 3.5 years of age predicts the complexity and sophistication of moral reasoning at 5.5 years of age. Further analysis suggests that ToM and moral reasoning—deontic reasoning about obligations and permissions in particular—are further intertwined, as not only is the former constitutive of the latter but also vice versa. Based on empirical and conceptual reasons, Wellman and Miller (2008) argue that deontic reasoning is also integral to a competent use of ToM, as reasoning about others' mental states involves not only belief-desire considerations but also an understanding of their obligations and permissions as part of their mental states.

Studies on the mentalist aspects of moral reasoning show that human infants are not only equipped with an understanding of intentional mental representational states, as well as with a basic capacity for empathy and emotion understanding, but also that both these mentalist elements of human cognition are constitutive of moral cognition from early on. The evidence strongly supports the claim that ToM and emotion understanding are also part of what, following Hamlin and her colleagues, are called the *foundations* of human moral cognition. Given that they appear from early on in infancy, ToM and emotion understanding are considered to be prelinguistic capacities. Thus, there are good reasons to think that they are universal. All human beings are endowed with such capacities and all of them deploy such cognitive capacities as constitutive parts of moral cognition, barring abnormal cognition, of course.

The evidence on ToM and emotion understanding further supports the idea that moral competence is closely related to the very nature of human cognitive endowment. It strongly supports the idea that competent moral cognition and development is heavily mentalist, as it requires an understanding of intentions, mental representational states, and emotions from early on. The evidence further shows that these elements are, together with the moral

principles and social evaluation abilities previously described, universal across human individuals and part of an evolutionarily endowed foundation for moral cognition. This prelinguistic moral capacity appears to describe a rather composite cognitive domain, one that includes complex evaluative capacities, principles of moral cognition, mentalist understanding of agents, empathy and affective understanding, and the resulting interactions among them.

Personal and Cultural Variability

Now, even though all human beings possess the same elaborate, principle-based, mentalist, affective and empathic prelinguistic moral capacity, this capacity *develops* differently across individuals and populations. These developmental differences naturally result in a heterogeneous mosaic of moral competence among adults with different behavioral patterns, different moral principles being endorsed (and socially enforced), as well as different individual and social attitudes towards normative principles in general.

As is natural, developmental differences may be found already in young infancy. Parent-child relations vary from subject to subject. In some cases the relation is "close, mutually binding, cooperative and affectively positive." (Kochanska 2002) These are known as parent-child relations with "mutually responsiveness orientation" and are characterized by a shared positive affect and a mutual responsiveness between parent and child. The presence or absence of this kind of parent-child relation is heavily decisive, as it determines whether, when, and how the child develops an internal guidance system for behavioral regulation, commonly known as "moral conscience," which works independently of any direct external influence or control. A strong moral conscience is generally considered to be necessary for the child to be properly integrated into a broad network of values and norms that surpass those that may be part of the parent-child relation.

Developing a mutually responsive relationship during the first year of life is determinant of the appearance of a strong moral conscience in the second year of life, and maintaining such a relationship during the toddler years predicts the nature of moral conscience during preschool and early school years (see Kochanska 2002 for a review). Perhaps surprisingly, the mechanisms by means of which mutually responsive parent-child relations influence the development of moral conscience are rather simple. Mutually responsive relations foster positive emotions and happy moods in the child, who in turn shows a better disposition for prosocial behavior. But not only, these relations also foster more receptiveness to parental guidance. As such, mutually responsive relations constitute the means for a child to properly internalize paternal moral rules and behavioral control. When the child is not

offered a mutually responsive relation a considerably high degree of effortful control will be needed to properly internalize the relevant rules and develop a moral conscience (see Kochanska and Kim 2014). Effortful control may, in fact, be the only means for such a child to develop morally, whereas children benefited by a mutually responsive parent-child relation may rely on multiple alternative mechanisms (see Kochanska and Kim 2014; Kochanska, Koenig, Barry, Kim, and Yoon 2010).

Several longitudinal studies have shown there is a stable pattern of moral functioning from infancy into adulthood, both for prosocial and antisocial behavior. Eisenberg and colleagues (see Eisenberg, Spinrad, and Knafo-Noam 2015 for a review; see Eisenberg and Fabes 1998 for a meta-analysis) have found that prosocial behavior, moral reasoning, empathy, and perspective taking remain stable, with an increase during early school years, a decline during adolescence, and a rebound into adulthood. Parallel results have been found for antisocial behavior including irritability, hostile rumination, and moral disengagement (see Caprara 1986; Caprara, Paciello, Gerbino, and Cugini 2007). Thus, the individual differences in moral cognition that appear in infancy, both for prosocial (see Kochanska 2002) and antisocial behavior (see Caprara, Barbaranelli, Pastorelli, and Perugini 1994; Caprara, Alessandri, Fida, Tisak, Fontaine, and Paciello 2014), are typically maintained if not augmented in adulthood.

Additional individual differences in moral cognition appear to be related to a subject's personality and personal identity. Moral personality has been shown to mediate out-of-sight observance of maternal rules from 25 to 52 months of age and predicts adaptation at 80 months of age (see Kochanska, Koenig, Barry, Kim, and Yoon 2010). Moral personality, in turn, appears to be determined by the quality of the parent-child relation, parental scaffolding, and the role these elements play in the subject's autobiographical narrative (see Lapsley and Hill 2009). Studies with 15 to 18-year-old subjects suggest a strong connection between moral personality / identity and the ideal self, understood as determining the person that the subject wants to become. This moral ideal self, in turn, positively predicts prosocial behavior and negatively predicts aggressive behavior (see Hardy, Walker, Olsen, Woodbury, and Hickman 2014).

Now, aside from *individual* differences the mosaic of human moral competence importantly exhibits multiple variegated differences at a *cultural* level. It has been shown that, in general, culture-unique values, environments, and practices do shape moral competence in children (see Super and Harkness 1997 and 2002). In a twenty-year-long project studying socialization practices with infants and their effects in prosocial reasoning and behavior in six different cultures—Kenya, Japan, India, the Phillipines, Mexico, and the United States—the Whittings and their colleagues (see Whitting and Whitting

1975; Whitting and Edwards 1988) found that levels of prosocial behavior varied across cultures, with higher levels being found in cultures with larger families and with a greater importance for women. As expected, children were exposed to different levels and forms of prosocial behavior across cultures, resulting in differences in moral reasoning and judgment.

A series of comparative studies with Mexican Americans and European Americans show that the former prefer cooperative behavior to a greater extent than the latter (see Knight, Bernal, and Carlo 1995). Mexican Americans also intend to inculcate a sense of collectivism and a great moral relevance for the family while European Americans do not (see Ramirez and Castaneda 1974; Steward and Steward 1973). High-income Mexican American families valued more what others did to help them than high-income European American families (see Williams 1991). These differences are reflected in adolescent moral reasoning and judgment, with Mexican American children and adolescents exhibiting a greater preference for cooperative over competitive behavior than European Americans (see Eisenberg and Fabes 1998). Further studies have shown, more generally, that parental values and practices that value family identity and family ties predict pro-family social behavior in the U.S. Latino population (see Knight and Carlo 2012). Studies of other minority groups in the United States confirm the relevance of cultural structuring for moral development and prosocial behavior. A recent study with 241 U.S. Jewish 14- to 17-year-olds showed that group exclusion of outside members was considered acceptable in the community context but not in the peer context. Yet the degree of acceptability of group exclusion varied according to how much intergroup contact the subjects had (see Brenick and Killen 2014).

The variety of human moral reasoning, thought, and judgment is of course much more diverse than what I have here described. There is moral diversity across cultural groups, but cultural groups are themselves constituted by diverse individuals, and the same happens within families and other social clusters. This immense diversity is commonly used as evidence against moral realism (see Mackie 1977) as moral diversity is associated with some kind of intractable disagreement among diverse moral views. Such disagreement, it is argued, is importantly distinct from disagreement about factual matters (see Stevenson 1944 and 1963), giving us reason to believe there are no objective moral facts. For if there were such facts, we would be forced—it is alleged—to claim that some human groups have better epistemic access to moral facts than others, thus resolving the diversity and disagreement of moral views. Alternatively, the argument goes, we can accept that morality is relative to the "way of life" that each human group endorses (see Mackie 1977), or that morality is mainly about expressing one's emotions and getting others to act accordingly (see Stevenson 1944 and 1963). There are, of course, multiple replies from a diverse group of moral realists, some rejecting the claim that

there is so much cultural and group-based moral variation, others arguing that such variation can be explained from a realist viewpoint (see Brink 1984; Shafer-Landau 1994; Loeb 1998; Lillehammer 2004; and Doris and Plakias 2008). I believe, however, that we need an account that both endorses this immense variation while explaining *and* predicting it.

We can get such explanation by looking into moral cognition. If it can be shown that the very same cognitive capacity that is apt for developing into a mature, competent form of diverse and varied adult cognition is also capable of delivering moral knowledge, it will offer a stronger basis for a satisfactory account of moral diversity and, *a fortiori*, a more pleasing defense of naturalist moral realism. I will describe how this can be done in what remains of this chapter. I will try to show that moral knowledge is better understood in relation to moral competent cognition. According to the evidence just reviewed, this cognitive ability is the result of evolutionarily endowed prelinguistic moral capacities, including prelinguistic moral principles, social evaluation capabilities, mental representational and intentional understanding of others, as well as emotion understanding and early empathic attunement. From this perspective, moral knowledge is the result of moral development, itself a product of cognitive development and maturation and not so much a matter of factual and empirical research and discovery, at least not primarily so.

7.3 THE LANGUAGE ANALOGY

The empirical studies briefly presented deliver a substantial understanding of moral cognition, its nature and function. On this understanding, moral cognition is constituted by an evolutionarily endowed, unlearned cognitive apparatus changing along the developmental path through the lifespan. From early on in infancy, this cognitive apparatus includes evaluative abilities, prelinguistic moral principles of cognition, a context sensitive capacity to understand action as goal-directed, and an early affective capacity for empathy. Since this cognitive apparatus gives form to the thought and cognition of very young prelinguistic infants, it is fair to say that it is universal. Especially with respect to prelinguistic moral principles, no matter how many distinct ones turn out to be, all human infants have them. The evidence also suggests that moral cognition is heavily mentalist, given the presence of intention understanding, ToM, and emotion understanding in early infant moral cognition. As such, the mentalist elements of moral cognition are also universal across human subjects.

As a whole the evidence seems to describe an intricate higher order cognitive ability that recruits resources from multiple domains and benefits from their interaction. The evidence also presents further results that are of

consequence for the philosophical debate about the nature of morality. First, it is clear that morality is unlearned in that human infants do not require guidance to learn it and in that, even though they may learn morality through social interaction, they are already endowed with some of it, and with the necessary means to acquire more. Second, human subjects can easily acquire different moralities, or different moral codes. Children, in particular, may do so from other children, without the presence or instruction of an adult. Third, and last, when this moral cognitive apparatus is fully developed, as it typically is in adult moral cognition, subjects are capable of encountering countless new scenarios that they will recognize as morally relevant and demanding for moral reasoning and judgment.

These three features are reminiscent of what Chomsky (1959) famously points out about natural language. Chomsky (1959) objects that the behaviorist has a narrow understanding of the study of linguistic behavior. On this narrow conception, claims Chomsky, the goal is "to provide a way to predict and control verbal behavior by observing and manipulating the physical environment of the speaker." (Chomsky 1959, 26) Yet, continues Chomsky

> One would naturally expect that prediction of the behavior of a complex organism (or machine) would require, in addition to information about external stimulation, knowledge of the internal structure of the organism, the ways in which it processes input information and organizes its own behavior. These characteristics of the organism are in general a complicated product of inborn structure, the genetically determined course of maturation, and past experience. (Chomsky 1959, 27)

Chomsky presents further reasons to reject the narrow methodology of the behaviorist, such as the fact that young infants have an unlearned understanding of language; the fact that infants can easily acquire a complete understanding of any natural language, even without the meticulous care of adults; and the fact that, when this linguistic capacity is fully developed, as happens in adulthood, competent subjects have the wherewithal to encounter an unlimited number of new linguistic structures that they will nonetheless recognize and understand without hesitation. "These abilities indicate," Chomsky famously argues, "that there must be fundamental processes at work quite independently of 'feedback' from the environment." (Chomsky 1959, 42)

The evidence of human moral cognition shows that something similar to what Chomsky (1959) claims about natural language is true about human morality. This is true in both a methodological and a psychological sense. Methodologically speaking, as happens with natural language, if we want to properly understand morality and gain some predictive power, knowledge of the internal workings of the human mind is required. Psychologically,

or cognitively, speaking there is, as I will show, a strong analogy between language and morality, as they seem to have a cognitive architecture in common. And this language analogy is an illuminating one, as there seem to be no deep metaphysical, epistemological and semantic qualms with the idea of knowledge of language.

Let me now present the dimensions within which I claim that language and morality are cognitively analogous. To begin with, both linguistic and moral cognition are products of human *evolutionary cognitive endowment* and, consequently, both are already exercised in early infancy. There is prelinguistic preparedness for both, language and morality, including prelinguistic (grammatical and moral) principles that will help an infant acquire the language—or the moral code—of her peers.

Now, moral competence is analogous to linguistic competence also in terms of development. Moral development results from the interaction between several domain general cognitive mechanisms. Surprisingly, a good number of these appear to be central for both, language and morality, including the understanding of mental states, the ability to identify intentions, to participate in joint attention, and to use practical reasoning. Fundamental to both, moral and linguistic competence, are social interactions and a substantial mentalist understanding of others. And just as linguistic competence, moral competence is a product of higher order cognition, involving multiple cognitive mechanisms. As such, moral competence may be considered a supermodular cognitive ability in the same sense as natural language is said to be (see section 4.2). This close relation between moral and linguistic cognitive development is further substantiated by studies on abnormal cognition, both moral and linguistic. Studies on autism have shown that this cognitive impairment exhibits problems in prosociality and ToM (see Leslie 1987). Recent studies have confirmed that competence in ToM directly determines moral development, and that ToM, mediated by linguistic competence, determines moral development in autism (see Peterson, Slaughter, Moore, and Wellman 2016). Other studies have shown that competence in ToM is not only predictive of moral competence and development, but also of linguistic competence and development (see Bloom 2000; Sabbagh and Baldwin 2001). It seems that, at least in cases of autism, an impairment of ToM has serious effects for both linguistic and moral cognitive development (see Tager-Flusberg 2007).

Third, there is a similar degree of individual and socio-cultural variability in both moral and linguistic competence. As with morality, natural languages exhibit an ample and multi-dimensional variability. Individual differences in language acquisition and development have been studied profusely and appear pretty much in every area of language development (see Hoff and Shatz 2007 for a general overview). Sociocultural differences in linguistic

competence are self-evident. Not only are there obvious differences across different cultures—e.g., differences between Japanese, Spanish, Basque, Hungarian, and Mandarin—but there also differences within what may be ordinarily considered to be the same natural language—e.g., differences between Australian and U.S. English; differences between Mexican, Rioplatense, and Peninsular Spanish; etc. Mayan, English, and Basque differ radically from each other even though they result from the same human cognitive ability (or set of abilities). This, however, does not imply some kind of linguistic relativism precluding subjects from possessing a deep and common knowledge of language.

Sociocultural variations do not turn natural languages into something less natural. The same should be said, or so I contend, of sociocultural variations in moral cognition. There is some degree to which both, language and morality, exhibit something like Whorfian cultural relativity. Yet it is also clear that both are, to a very substantial and important degree, universal across humans. Different natural moralities—or distinct moral codes, if you prefer—appear to be the product of social conventions. But so are natural languages to an important degree (see Lewis 1989). This makes natural languages (at least partly) a population-relative human product, but it does not render them futile. Similarly, moral codes (those one can find across human populations) appear to be partly the result of abiding by certain conventions. This makes them (at least partly) a population-relative human product, but it does not render them useless or mistaken. And, lastly, as with natural languages, the conventionality of moral codes does not preclude any specific morality, or moral code associated to a given population, from being based on universal principles common to all other moral codes. In both cases, morality and language, we can find shared prelinguistic (perhaps innate) principles that partially determine how they will develop into a full-blown competence and give shape to the moral code, or natural language, they will end up acquiring and sustaining.

Fourth, moral and linguistic cognition are normative in a similar sense. They are both principle-based forms of cognition. The principles, upon which they are based, both linguistic and moral, may be of different sorts. They may be unlearned, and so fully universal, or acquired and even population-relative. They may be fully general and abstract—and so fully grammatical or purely moral—or constitute a more frugal and context sensitive heuristic set (see section 4.3; see also Gigerenzer 2008b). Yet all such principles are of central importance to both linguistic and moral cognition, as they help perform an essential cognitive function, namely, that of eliciting judgments of acceptability / unacceptability about their subject matter—whether linguistic or moral—and they have a direct say on the sort of behavior that corresponds to such judgments.

Aside from these four substantial cognitive similarities between moral and linguistic cognition, there are others that correspond to how competence is viewed by the subjects themselves. They are worth mentioning insofar as they reflect a common underlying mechanism. First, it is not common to see subjects correcting each other's moral or linguistic behavior. Subjects are simply not expected to fail—they are not expected to say something linguistically unacceptable, or to draw a morally unacceptable judgment. Second, when they exist, most corrections are directed to infants and young children, whose moral and linguistic cognition is still far from fully developed. Third, if addressed to what is taken to be a normal and competent adult, corrections are usually felt as inappropriate and perhaps even diminishing. Fourth, whenever it happens, moral or linguistic failure is commonly taken to be an exception related in some sense with incompetence, not with factual ignorance—though, of course, there may be mistakes that result from ignoring certain factual information, in which case they are not considered to be moral or linguistic.

Finally, there are two potential, yet non-troubling dissimilarities between language and morality from the point of view of competent subjects. First, some seem to think or have the impression, that there is more ordinary moral talk and discourse than any corresponding linguistic phenomena. Engaging in discourse about language presupposes specialized theoretical education, and this is not the case with moral discourse. Apparently, the impression goes, people ordinarily talk about morality even if they have not received any sort of theoretical education about it. Second, it is not surprising that there are gaps between competence and performance. Subjects may competently know a language and still fail to perform linguistically (for any sort of reason). If morality is like language, then we should also find such competence-performance gaps. Intuitively, however, it seems as if there is a substantial difference between moral and linguistic failures to perform. Moral gaps between competence and performance seem to be serious trouble, whereas linguistic ones are not. These alleged dissimilarities from within competent subjects' views about morality and language have, to my mind, a very simple and unproblematic explanation, namely, that morality matters more for ordinary life than language. Ordinary talk is usually about what matters most in ordinary life; and moral behavior is much more important for our practical everyday life—even more, it is more important for the species' survival—than linguistic behavior.

Language and morality are, I have argued, substantially similar. Like linguistic competence, moral competence is widespread and universal among humans. Human infants already exhibit moral understanding, suggesting that moral competence does not demand any kind of explicit education or learning. And moral competence, like linguistic competence, bestows competent

subjects with the needed equipment to identify, understand, and respond to an unlimited set of new morally relevant scenarios. Indeed, once we look at their cognitive nature, morality and language appear to be even closer. They have similar initial conditions, as they both have an evolutionarily endowed apparatus, including higher order cognitive abilities such as ToM, and intention understanding, as well as prelinguistic principles of cognition for the linguistic and moral domains. They have similar developmental trajectories, benefitting from substantial interaction among multiple domain general mechanisms (in both cases), as well as from social interaction. They both give place to an enormous and multi-dimensional set of individual and sociocultural variations. And, lastly, they both have an important normative function to perform, for which they carry their own set of principles.

I believe this is good news for the naturalist moral realist who thinks there is moral knowledge and the corresponding moral truths, and that the latter are warranted by moral facts and properties of a naturalist kind, which enjoy a mind-independent existence. If the linguistic analogy with morality is as substantial and strong as I have described, then we can only benefit from viewing morality from a point of view that approaches our best theorizing about natural language. Among students of language, no one seriously doubts that there is linguistic knowledge, and that it is objective even though natural languages differ substantially across sociocultural groups and even across individuals. And, finally, no theorist seriously doubts the naturalness of Spanish, English, Japanese, or Mandarin, or that each one of these gives place to its own set of linguistic truths. Languages, like morality, appear too early in human ontogeny for them to be considered cultural or social products that result from embracing a certain way of life. The language analogy illuminates our philosophical understanding of morality, suggesting an account of the latter that may avoid the philosophical problems described at the beginning of this chapter. In what follows I will present such an account and describe how it may solve or avoid the said problems.

7.4 OBJECTIVITY, CONVENTIONALITY, UNIVERSALITY, AND NORMATIVITY

At first glance there seems to be a tension between the *naturalist realism* and the *psychological* components of the account I am presenting. The former carries a commitment to a substantially *non-subjective* nature of moral truths and judgments, or at least some of them. Yet, the deep psychological nature of moral knowledge appears to go against this. If morality is, as I have been suggesting, substantially determined by the nature and state of human

cognitive endowment, then how is it that moral judgments may be said to be objective? If moral knowledge is *on a par* with knowledge of language, as it even includes a substantial set of culturally variable and conventional truths, in what sense is there moral objectivity or universality? I believe the apparent tension between a naturalist moral realism and the deeply psychological nature of morality, as described by the language analogy, is just that, an appearance. In this section I will show how the proposed *cognitive* version of naturalist moral realism is genuinely naturalist and realist by looking into the objectivity, conventionality, and universality of the resulting moral judgments.

Objectivity

To claim that moral facts and properties are natural facts and properties as postulated by empirical science does not imply that the corresponding facts and properties are objective in exactly the same sense in which a physical fact or property may be said to be objective. There is an important and substantial notion of *moral objectivity* that successfully applies to the set of moral facts and properties proposed.

Street (2006) famously argues against the possibility of having both, naturalist and objective moral facts and properties. Either normative facts are naturalist or they are objective, but they cannot be both. This is so because for them to be naturalist, they must have been the result of non-normative evolutionary processes. Yet to find out which are the normative facts that result from such non-normative processes we must use our evaluative judgment—i.e., we must use moral theory to determine which are the moral facts. If so, then the normative facts that will be so identified will not be *independent* from our evaluative judgments, and so they will not be objective. Street's (2006) dilemma poses no threat to the view I am here suggesting since the moral facts and properties *are* the result of *normative* evolutionary processes, namely, those concerned with limiting human cognition about human behavior. Normative truths and facts are facts about a biologically endowed human moral cognition. These are not relational moral properties, and there is no need for idealized and fully informed evaluative judgment to identify them; we only need careful empirical testing to identify the complete set of principles shaping moral cognition, the nature of moral competence and how these help us attain moral knowledge. Thus conceived, moral principles (e.g., prelinguistic ones) are completely *independent* of which moral views the experimenter or theorist holds. Hamlin and her colleagues (see Hamlin 2013) have successfully identified them and not by means of their evaluative judgment. As such, prelinguistic moral principles are objective in Street's robust sense.

Moral Objectivity: moral principles "hold independently of the *whole set* of evaluative judgments we make or might make upon reflection, or independently of *the whole set* of other evaluative attitudes we hold or might hold upon reflection." (Street 2006, 111)

The language analogy suggests there are various different types of moral principles, some of them being universal while others seem to be culture-relative. An especial group among the universal moral principles is the set of moral principles that are a central part of human moral cognition by setting the limits of such a especial form of cognition. Chief among such principles are those I have called "prelinguistic moral principles." It should be clear that such moral principles observe the above "robust" notion of moral objectivity, and the same goes for other moral principles that may result from properly exercising moral competence, for example by being justified by robustly objective prelinguistic moral principles.

Street's objection is based on the assumption that "clearly it is implausible to think that the acceptance of a *full-fledged evaluative judgment* with a given content—for example, the acceptance of the judgment that 'one ought to help those who help you'—is a genetically heritable trait." (Street 2006, 118–19) However, the empirical evidence directly falsifies this assumption. As Hamlin, Ullman, Tenenbaum, Goodman, and Baker (2013) show, very young prelinguistic infants already show an understanding and endorsement of some such principles, precisely of the kind and content that Street presents. Infants as young as 10 months of age infants already distinguish between helpers and hinderers and prefer the former to the latter, and unless this is the result of a genetically heritable trait, there is no explanation of how this may be the case.

Moral objectivity, as moral cognition, is of a special kind. There is no need for it to be represented propositionally, as part of the content of certain moral judgments or concepts. Moral objectivity is not a matter of *what* is morally claimed, but a matter of *how* such claims are interweaved in a context that renders their justification. If such claims are justified independently of the subject's normative attitudes, then they may be said to be objective. If they happen to be justified in virtue of being picked by natural selection, then they may be both, objective and universal.

Conventionality

Moral truths and moral principles are meant to solve coordination problems. Following Lewis (see Lewis 1969 and 1975) it may be said that a given solution to a coordination problem is arbitrary if, first, there is at least one other alternative solution; second, this other solution is equally well-suited to solve the problem; and, third, there is no good reason to prefer one solution instead

of the other. So, for example, driving on the right-hand side of the road is an arbitrary solution to the problem of avoiding head-on collisions. There is at least one alternative—i.e., driving on the left-hand side of the road—that is equally well suited to solve the problem and there is no good reason to prefer one solution instead of the other. On this account of arbitrariness, some moral truths (e.g., prelinguistic ones) turn out *not* to be arbitrary, whereas others (e.g., culture-relative ones) may be.

Prelinguistic moral principles, those shaping human moral cognition, are not arbitrary because there seem to be no alternative solutions that are equally well suited to solve the problem. Those principles are part of our cognitive biological endowment precisely because they were the best such solution—given human limitations, of course. Furthermore, they are uniquely well justified, as they are the ones resulting from natural selection. Culture-relative moral truths are not as uniquely well justified and their variety across human groups strongly suggests that there are a great number of equally well-suited and justified alternatives. This, however, requires careful empirical study, as it is highly probable that a great number of culture-relative moral principles may result from an attempt to solve a problem that is unique to a given population or environment and, in this sense, may seem to be (morally at least) less arbitrary.

It is important to point out that, whether or not they are arbitrary in virtue of being conventional, some moral truths or principles may still be objective in the robust sense above described. They may be offered as a solution to a coordination problem, they may be arbitrary solutions, and still it may be that these solutions hold independently of any evaluative attitude of the relevant population—e.g., if they are our best solution to a unique coordination problem faced by this or that particular population. Thus, it seems possible (and perhaps desirable) to distinguish between objective and non-objective moral *conventions*.

These are notions of moral objectivity and conventionality that a naturalist realist can accept. Objective moral truths—whether they are evolutionarily selected and developmental or just an arbitrary solution proposed by a given population to solve a specific coordination problem—are those that hold independently of any normative attitudes the relevant subjects may have. This notion salvages everything that is needed to account for the objectivity of moral practice. It is not the usual notion of objectivity associated with basic science, but that is a virtue of the theory since it does not impose a non-moral notion of objectivity.

Universality

If I am correct, then the conventionality of certain moral truths does not go against their objectivity. Yet, it may go against their universality, since they

are meant to be solutions to specific coordination problems faced by particular human populations. Still, there are certain moral truths that may be said to be universal in a plain and simple sense. Such are the moral principles that all human beings have in virtue of being cognitively competent human individuals. Some of these principles—i.e., prelinguistic moral principles—are part of our genetically transmitted cognitive endowment. Others result from a full development of such cognitive competence.

As a whole, universal moral principles, whether prelinguistic or developmental, are to be found through careful empirical testing. The only way to identify them is by carefully looking into human, whether infant, adolescent, or adult, moral cognition. Such principles are properly viewed as universal and objective. They are universal because all competent humans have them, and cannot opt out of them. They are objective because they are justified independently of what any human individual, or group of individuals, believe or endorse with respect to them. They are well justified because they are, directly or indirectly, the result of natural selection.

Normativity

If, as the theory claims, such principles shape human moral cognition, then they also exhibit another important moral feature, namely, they exercise behavioral control upon us, as is expected given their evolutionary relevance.

Moral principles of cognition *are* normative just as other principles of cognition are, they regulate the way in which the relevant task is to be performed. Linguistic principles regulate the way in which sentences are produced and processed; and visual principles regulate the way in which photonic stimuli is to be processed. Similarly, moral principles determine the way in which intentional goal-directed behavior is to be performed and evaluated. The difference between moral and other cognitive principles lies chiefly in the different tasks they are meant to regulate. In performing their regulatory cognitive function moral principles of cognition distinguish between acceptable / unacceptable, commendable / reproachable, praiseworthy / despicable behavior, and elicit the corresponding judgments and accompanying positive or negative motivation. As such, principles of cognition *in general* are both *natural* and *normative*. They are required by our best empirical theories to meet their explanatory goals (e.g., account for human vision), and they play a normative, regulatory function. Moral principles of cognition are no exception. Contrary to what recent views defend (see Joyce 2006), satisfactory evolutionary accounts of morality do not make the normativity of moral principles of cognition redundant. Rather, they are necessary if we want to properly account for the evidence (see section 7.2).

7.5 KNOWLEDGE, MOTIVATION, AND SEMANTICS

If morality, like language, is determined by the nature and state of human cognitive endowment in what sense is there knowledge of morality? Which are the moral truths and how can they be known? What is the meaning of moral terms and how are complex moral statements interpreted? And finally, how can knowledge of these truths have any motivational force? Let me try to answer these questions one by one.

Knowledge

In what sense is there knowledge of morality? Following the analogy with language it seems natural to say that knowledge of morality is better understood in relation to moral competence. Like knowledge of language, moral knowledge is *both* the capacity to exercise competent moral cognition—i.e., a complex set of abilities including emotion understanding, ToM, social evaluation, and reasoning under moral principles—and the knowledge obtained through such exercise. Subjects are said to know morality in virtue of possessing such principle-based cognitive ability as part of their moral cognitive machinery and, thus, being capable of understanding (and perhaps discovering) further moral claims and truths. Subjects need not know *that* they possess such moral cognitive endowment in order to be competent moral subjects. Competent subjects simply know how to engage in moral thinking and acting.

Since Ryle (1946) it is common to distinguish between two kinds of knowledge: *knowledge that* and *knowledge how*. The former constitutes knowledge of propositions, things that can be true or false—e.g., knowing that helping others is the right thing to do. The latter is rather concerned with how to do something—e.g., knowing how to morally evaluate a certain situation. As such, on the view I am presenting, knowledge of morality seems to be a kind of *knowledge how*. Most realist views of morality, however, assume that moral knowledge is a kind of *knowledge that*, since there are moral truths to be known (see Kitcher 2011, for a know-how account of moral knowledge).[1]

Yet, even though the very idea of competence seems closer to some kind of knowledge how, the specific nature of *moral* competence cannot be reduced to just another instance of knowledge how. Moral competence includes not only cognitive abilities but also informational / representational states chiefly constituted by principles of moral cognition (see Hamlin 2013; Geraci and Surian 2011). As such moral competence also serves as the foundation for *propositional* moral knowledge, as the principles of moral cognition may ground the truth of some or other proposition. For example, according to this view young infants *know how* to distribute scarce resources among parties, because they are endowed with a prelinguistic principle of fairness (see Geraci and Surian

2011). In virtue of the latter it may also be said that 12-month-olds *know that* equal distributions are the right ones. Surely, we may (and do) attribute implicit propositional knowledge to competent subjects; but besides this point, adult subjects may, given enough reflection and theoretical training, come to know it explicitly. Thus, even though moral competence is a kind of *know how*, there is also a corresponding *know that* (see Bengson and Moffet 2012, for a review of the debate on the *know how—know that* relation).

Which are the moral truths? Given that for each moral judgment there is a corresponding truth-evaluable proposition, by claiming that moral truths are directly or indirectly grounded or justified by the principles of moral cognition, this view is open to the existence of different kinds of moral truths, at least as many as there are kinds of moral principles to justify them. Moral truths that are directly justified by objective and universal moral principles will seem to be fully objective and universal moral truths. Those that are only supported by conventional, culture-relative principles may be considered as culture-relative moral truths. And, of course, there may be many moral truths of different sorts in between. Naturally, some kind of consistency between universal and culture-relative truths should be expected since it is by means of the same cognitive ability that all moral truths come to be known. There is a basic "innate moral core" (as Hamlin 2013 puts it) constituting such cognitive capacity, a set of foundational moral truths that are not negotiable. Perhaps there are other principles of moral cognition that also play a foundational or structural role. This can only be determined through empirical research.

What is the metaphysical nature of these truths? Moral facts, facts about what is right and wrong, are determined by the nature of human moral cognition, by the fundamental moral principles that shape such cognitive processing. Human moral cognition, like the rest of human cognition, is as natural as any other biologically endowed human trait genetically bestowed from one generation to the other. Once again, moral cognition is not metaphysically distinct from linguistic or visual cognition in humans. They are all metaphysically on a par. This cognitive or psychological nature of moral facts does not make them less suitable for a moral realist theory. Cognitive-psychological moral truths and the principles justifying them meet Railton's (1986) *reality* criteria of independence and feedback (see section 7.1).

> *Independence*: it exists and has certain determinate features independent of whether we think it exists or has those features, independent, even, of whether we have good reasons to think this;
>
> *Feedback*: it is such—and we are such—that we are able to interact with it, and this interaction exerts the relevant sort of shaping influence or control upon our perceptions, thought, and action. (Railton 1986, 172)

Moral principles of cognition are whatever it is that they are as a result of evolutionary processes on human cognition. If they were evolutionarily selected it is because these principles, and the role they play with the other elements of moral cognition, have exerted an indispensable role in shaping and influencing human thought, perception and, of course behavior in ways that have contributed to the survival of the species. Thus, these principles exist with all their features independently of whether we think they exist, and have whichever features they have independently of whether we have good reasons to think they have them. Moral truths and the moral principles of cognition upon which they stand are very real and very natural metaphysically speaking (see Kumar 2015).

How can these truths be known? A detailed answer to this question can only be drawn from a detailed and long review of empirical studies on moral cognition and development. For a brief and partial look at it, go back to section 7.2. A short answer should be clear by now. Human infants are born with a biologically endowed moral cognitive apparatus. Prelinguistic moral principles are an integral part of it. Thus, we know these truths simply because we are born with them or, if you prefer, we are born with the capacity to know them, and will do so with enough reflection and development of our cognitive apparatus in general. Other moral truths are not evolutionarily delivered to us. Some of these truths may be objective—and perhaps even universal across humans, not evolutionarily but developmentally so—while some other may merely be culture-relative. All of them are knowable as a result of the exercise and development of our moral cognitive apparatus, including its interaction with cognitive mechanisms and abilities of a general domain—i.e, practical reasoning, statistical analysis, and general knowledge.

Motivation

There are two remaining problems that have proven to be unmanageable for naturalist moral realist accounts (see section 7.1). These are the semantic and motivational problems. I will conclude this chapter by showing how the proposed view avoids such problems. Consider first the motivational problem. If knowledge of moral truths is knowledge of natural properties, how can it have any motivational force? Naturalist realist accounts of moral truths usually have problems answering this question precisely because moral facts are taken to be *non-psychological*. The existence and characteristics of moral facts so conceived has no obvious bearing on a subject's psychology. Even if they are meant to have relational, dispositional properties (see Railton 1986; and Lewis 1989) it is, at best, merely assumed that non-psychological facts and properties may have such an extraordinary psychological effect, without any satisfactory explanation of why and how that might be so. Without

the help of some internally motivating state—it is argued following Hume (2007)—there is no reason to think that knowledge of an external state of affairs will somehow exert a motivational force.

It should be clear that these problems do not arise for the naturalist moral realism here sketched. Moral facts, to begin with, are *cognitive-psychological* facts. Knowing them is, first and foremost, a matter of having a certain competence, moral competence, to perform in such and such a way—i.e., to evaluate certain scenarios, and to respond to certain judgments. As the evidence shows (see section 7.2), moral competence makes use of a complex set of psychological abilities, including emotion understanding and affective attunement. Thus, moral knowledge seems to be already accompanied by a host of internal, psychological, and intrinsically motivating mental states. Being morally competent is also a matter of having the appropriate emotional response to a given scenario. Being motivated or unmotivated seems to be another feature of moral cognition. This account of the motivational force of moral knowledge and moral judgment is fully compatible with the idea that moral facts are attitude-independent yet internal psychological entities.

Semantics

Finally, what is the meaning of moral terms such as "good," "right," and "wrong?" More specifically, how does this account avoid the "open question" objection against naturalist moral realism? Naturalist and realist views of moral discourse tend to have problems with the meaning of moral terms because the properties these terms are taken to denote are typically considered to be non-normative *in virtue of* being natural. If so, then it follows from these views that normative terms such as "good," "right," or "wrong," turn out to denote non-normative properties, such as *being approved of by ideal social rationality* (see Railton 1986). Non-normative accounts of the meaning of normative terms inevitably stumble against Moore's open question arguments. Moore (1903) famously points out that if such accounts of meaning for moral terms were correct, it would be odd to ask something of the form "I know that doing such and such is F, but is it good?" where "F" stands for the naturalist non-normative property allegedly denoted by moral terms. The problem is that such questions do not seem to be odd. As Railton (1986) recognizes, it does not seem odd to ask "I know that doing such and such is approved of by ideal social rationality, but is it good?" (see Railton 1986, 204–207).

On the view I am here describing, normative or moral terms *do not* denote non-normative properties. Moral facts and properties are both *natural and normative* in virtue of being cognitive psychological entities, more precisely, principles of moral cognition. These principles *determine* what is to be deemed good, right, wrong, praiseworthy, despicable, etc. Thus, we can keep an ordinary, unanalyzed meaning of, say, "good" and claim, on a theoretical

yet non-semantic level, that performing this or that goal-directed action is good if it is so determined by the relevant principles of competent moral cognition. It certainly seems odd to ask "I know that doing such and such a thing is good, but is it good?"

How is it that competent *speakers* manage to denote such psychological properties? According to open compositionality competent speakers acquire language through multiple different mechanisms and strategies involving a host of different domains of cognition. Language is viewed as a supermodular and highly interactive cognitive capacity. There is, of course, reason to think that moral cognition and moral competence is among the cognitive abilities with which such supermodular capacity interacts. If, as the proposed view claims, competent subjects know what is right and what is wrong in much the same way in which competent speakers know what is grammatical and what is not, there is little to be said about how speakers manage to use moral terms successfully to denote the relevant psychological properties. Subjects, so to speak, are already granted with the relevant moral concept, given that they are cognitively competent. There is no need for subjects to establish the proper mapping between "good" and the moral property it aims to denote. That mapping is already engrained as part of their biologically endowed cognitive architecture.

In section 7.1, I claimed that Moore's open question argument was based on the assumption that language observes *strong compositionality*, thus creating trouble for the naturalist realist when it comes to explaining how "good" makes its contribution to the meaning of any complex expression (such as a question). Once we give up the closed view and its endorsement of *strong compositionality*, and instead embrace *open compositionality* (see sections 4.3 and 4.5) we need not offer a compositional, syntactically driven, interpretation of complex expressions involving moral terms. Determining the meaning of any complex expression is always the result of a decision-making process aimed at finding the most *appropriate* and *economical* interpretation given *limited* information. This process may be the result of applying syntactic and semantic considerations or, instead, it may result from following heuristic strategies for fast and frugal decision-making. Given that speakers are equipped with a specialized and highly complex moral cognitive architecture, it seems natural to expect that all complex expressions involving moral terms are processed primarily by such apparatus, with linguistic considerations—whichever they may be—playing only a secondary role. How exactly this happens is, naturally, an empirical question in need of proper testing.

In this chapter I have tried to offer a mere picture of what moral knowledge would look like by following the *cognition-first* methodology (see section 4.4). The empirical studies on human moral cognitive development strongly suggest that moral knowledge is determined by the nature and state of human cognitive endowment. I have also argued that there is a strong analogy between language and morality cognitively speaking. Like linguistic

competence, moral competence is widespread and universal among humans, appearing from early on in infancy. Moral competence, like linguistic competence, bestows competent subjects with the tools to understand and respond to an ever-changing moral scenario. Language and morality exhibit parallel initial conditions in infancy and also similar developmental trajectories. They recruit an extraordinarily similar supermodular cognitive architecture. They both give place to an enormously diverse set of individual and sociocultural variations. And they both have an important normative function to perform, for which they carry their own set of principles. This analogy strongly suggests that what is special about morality is moral cognition, not the existence of *sui generis* moral facts or properties. To properly account for the evidence we need to postulate a dedicated set of cognitive resources aimed at processing information and guiding behavior. Moral knowledge is, on this view, a consequence of competently using this cognitive capacity. Moral objectivity is a result of moral evolution (human cognitive evolution), and moral variability is a consequence of moral development. Like natural language, morality will flourish differently across different populations. I have tried to describe how such a view of moral knowledge would look like. What I have offered is little more than a picture, the blueprint of an account. A fully developed theory is certainly desirable, but I am afraid it would require a book of its own.

NOTE

1. It is important to distinguish the present proposal from what Kitcher (2011) dubs "pragmatic naturalism," the view that moral knowledge is a kind of *know how* with no associated *know that*—there are no moral truths on this view—which humans have inherited through some kind of cultural descent—on Kitcher's view "our remote ancestors (. . .) *invented* ethics." (Kitcher 2011, 3) The view here proposed differs radically from pragmatic naturalism. There are moral truths and, thus, there is moral *know that;* and nobody invented morality, even our remote ancestors had it already as they were young infants. Morality *is* transmitted through biological descent; cultural transmittance is made possible thanks to it.

Bibliography

Adams, F., and Stecker, R. 1994. "Vacuous singular terms." *Mind & Language*, 9, 4, 387–401.

Anscombe, G. E. M. 1957. *Intention*. Oxford: Basil Blackwell.

Akhtar, N. 2005. "The robustness of learning through overhearing." *Developmental Science*, 8, 199–209.

Akhtar, N., and Tomasello, M. 2000. "The social nature of words and word learning." In *Becoming a Word Learner: A Debate on Lexical Acquisition* edited by R. M. Golinkoff and K. Hirsh-Pasek, 115–35, Oxford: Oxford University Press.

Akhtar, N., Callanan, M., Pullum, B., and Scholz, G. 2004. "Learning antecedents for anaphoric one." *Cognition*, 93, 141–45.

Ambridge, B., and Lieven, E. V. M. 2011. *Child Language Acquisition. Contrasting Theoretical Approaches*. Cambridge: Cambridge University Press.

Aslin, R. N., and Pisoni, D. B., 1980. "Some developmental processes in speech perception." *Child Phonology*, 2, 67–96.

Aslin, R. N., Pisoni, D. B., Hennessey, B. L., and Perey, A. J. 1981. "Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience." *Child Development*, 52, 4, 1135–45.

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

Bach, K., 2005. "Context *ex machina*" In *Semantics versus Pragmatics* edited by Z. Szabó, 15–44, NewYork: Oxford University Press.

Baggio, G., van Lambalgen, M., and Hagoort, P. 2012. "The processing consequences of compositionality." In *The Oxford Handbook of Compositionality* edited by M. Werning, W. Hinzen and E. Machery, 554–672. Oxford: Oxford Univesity Press.

Bailey, K. G. D., and Ferreira, F. 2003. "Disfluencies affect the parsing of garden-path sentences." *Journal of Memory and Language*, 49, 183–200.

Baldwin, D. A. 1991. "Infants' contribution to the achievement of joint reference." *Child Development*, 62, 875–90.

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., and Tidball, G. 1996: "Infant's reliance on a social criterion for establishing word-object relations." *Child Development* 67, 3135–53.

Baldwin, D. A., and Meyer, M. 2007. "How inherently social is language?" In *Blackwell Handbook of Language Development* edited by E. Hoff and M. Shatz, 87–106. Oxford: Blackwell.

Baldwin, D., and Moses, L. J. 2001. "Links between social understanding and early word learning: Challenges to current accounts. *Social Development*,10, 309–29.

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

Barton, S. B., and A. J. Sanford. 1993. "A case study of anomaly detection: Shallow semantic processing and cohesion establishment." *Memory and Cognition*, 21, 477–87.

Bengson, J., and Moffet, MA. 2012. "Two conceptions of mind and action: Knowing how and the philosophical theory of intelligence." In *Knowing How: Essays on Knowledge, Mind, and Action* edited by J. Bengson and M. A. Moffet, 3–58, Oxford: Oxford University Press.

Bever, T. G. 1970. "The cognitive basis for linguistic structures." In *Cognition and Language Development* edited by R. Hayes, 279–362, New York: Wiley and Sons.

Bloom, L. 1970. *Language Development: Form and Function in Emerging Grammars.* Cambridge, MA: MIT Press.

Bloom, P. 1990. "Syntactic distinctions in child language." *Journal of Child Language*, 17, 343–55.

———, 1993. "Overview: Controversies in Language Acquisition." In P. Bloom (ed.) *Language Acquisition: Core Readings*. New York: Harvester.

———, 1998. "Language acquisition in its developmental context." In *Handbook of Child Psychology: Vol. 2: Cognition, Perception and Language* edited by W. Damon (series ed.) and D. Kuhn and R. S. Siegler (vol. eds.), 309–70, 5th edition, New York: Wiley.

———, 2000. *How Children Learn the Meaning of Words*. Cambridge, MA: MIT Press.

Bornstein, M. H. 1987. "Perceptual categories in vision and audition." In *Categorical Perception: The Groundwork of Cognition* edited by S. Harnad, 287–300, Cambridge University Press.

Bornstein, M. H., Kessen, W., and Weiskopf, S. 1976. "The Categories of Hue in Infancy." *Science*, 191, 201–2.

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.

Brandom, R. 1994. *Making it Explicit*. Cambridge, MA: Harvard University Press.

Brandt, R. 1979. *A Theory of the Good and the Right*. Oxford: Clarendon Press.

Braun, D. 2005: "Empty names, fictional names, mythical names." *Nous*, 39:4, 596–631

Brenick, A., Killen, M. 2014. "Moral judgments about Jewish-Arab intergroup exclusion: The role of cultural identity and contact." *Developmental Psychology*, 50, 1, 86–99.

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

Brink, D. O. 1984. "Moral realism and the skeptical arguments from disagreement and queerness." *Australasian Journal of Philosophy*, 62, 2, 111–25.

Brink, D. O. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Brock, S. 2004. "The ubiquitous problem of empty names." *Journal of Philosophy*, 101, 6, 277–98.

Brown, R. 1973. *A First Language: The Early Stages*. London: George Allen & Unwin.

Brown, R., and Hanlon, C. 1970. "Derivational complexity and the order of acquisition in child speech." In *Cognition and The Development of Language* edited by J. R. Hayes, 11–54, New York: Wiley.Bruner, J. S. 1975. "The ontogenesis of speech acts." *Journal of Child Language*, 2, 1–19.

Brugman, D., Keller, M., and Sokol, B. 2013. "Introduction: Meaning, measurement, and correlates of moral development [Special issue]." *European Journal of Developmental Psychology*, 10, 2, 99–105

Bruner, J. S. 1983. *Child's Talk: Learning to Use Language*. New York: W. W. Norton.

Burnett, H. 2017. *Gradability in Natural Language. Logical and Grammatical Foundations*. Oxford: Oxford University Press.

Bybee, J., 1985. *Morphology*. Amsterdam: Benjamins.

Byrne, A. 2009. "Experience and content." *The Philosophical Quarterly*, 59, 236, 429–51.

Byrne, R. M. J. 2005. *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge, MA: MIT Press.

———. 2016. "Counterfactual Thought." *Annual Review of Psychology*, 67, 135–57.

Camblin, C. C., Gordon, P. C., and Swaab, T. Y. 2007. "The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking." *Journal of Memory & Language*, 56, 103–28.

Campbell, R. 2007. "What is moral judgment?" *The Journal of Philosophy*, 104, 7, 321–49.

———. 2014. "Reflective equilibrium and moral consistency reasoning." *Australasian Journal of Philosophy*, 92, 3, 433–51.

Caprara, G. V. 1986. "Indicators of aggression: The dissipation-rumination scale." *Personality and Individual Differences*, 7, 6, 763–69.

Caprara, G. V., Alessandri, G., Fida, R., Tisak, M. S., Fontaine, R. G., Paciello, M., 2014. "The contribution of moral disengagement in mediating individual tendencies toward aggression and violence." *Developmental Psychology*, 50, 1, 71–85.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., and Perugini, M. 1994. "Individual differences in the study of human aggression." *Aggressive Behavior*, 20, 4, 291–303.

Caprara, G. V., Paciello, M., Gerbino, M., and Cugini, C. 2007. "Individual differences conducive to aggression and violence: Trajectories and correlates of irritability and hostile rumination through adolescence." *Aggressive Behavior*, 33, 4, 359–74.

Carey, S. 1997. "Do constraints on word meanings reflect prelinguistic cognitive architecture?" *Cognitive Studies. Bulletin of the Japanese Cognitive Science Society*, 4, 1, 35–57.

———. 2001. "Whorf vs. Continuity theorists: Bringing data to bear on the debate." In *Language Acquisition and Conceptual Development* edited by M. Bowerman and S. C. Levinson, 185–214. Cambridge: Cambridge University Press.

Carlo, G. 2006. "Care-based and altruistically based morality." In *Handbook of Moral Development* edited by M. Killen and J. G. Smetana, 551–79, Mahwah, NJ: Erlbaum.

Carnap, R. 1947. *Meaning and Necessity. A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

Chalmers, D. J. 2002. "On sense and intension." *Philosophical Perspectives, 16, Mind and Language*, 135–82.

Chierchia, G. 2013. *Logic in Grammar. Polarity, Free Choice, and Intervention*. Oxford: Oxford University Press.

Chomsky, N. 1959. "A review of B. F. Skinner's *Verbal Behavior*." *Language*, 35, 1, 26–58.

———. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

———. 1980. *Rules and Representations*. London: Basil Blackwell.

———. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

———. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Chouinard, M., and Clark, E. 2003. "Adult reformulations of child errors as negative evidence." *Journal of Child Language*, 30, 637–69.

Christianson, K., Hollingworth, A., Halliwell, J., and Ferreira, F., 2001. "Thematic roles assigned along the garden path linger." *Cognitive Psychology*, 42, 368–407.

Christianson, K., Williams, C., Zacks, R., and Ferreira, F. 2006. "Younger and older adults' good enough interpretations of garden-path sentences." *Discourse Processes*, 42, 205–38.

Church, A. 1941. *The Calculus of Lambda Conversion*. Princeton: Princeton University Press.

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

Copp, D. 1995. *Morality, Normativity, and Society*. New York: Oxford University Press.

Cushman, F., Young, L., and Hauser, M. 2006. "The role of conscious reasoning and intuition in moral judgment." *Psychological Science*, 17, 1082–89.

Cutting, J. E., and Rosner, B. S. 1975. "Categories and boundaries in speech and music." *Perception and Psychophysics*, 16, 564–70.

Dancy, J. 1993. *Moral Reasons*. Cambridge: Wiley Blackwell.

Darwall, S., Gibbard, A., and Railton, P. 1997. *Moral Discourse & Practice: Some Philosophical Approaches*. Oxford: Oxford University Press.

Davidov, M., Zahn-Waxler, C., Roth-Hanania, R., and Knafo, A. 2013. "Concern for others in the first year of life: Theory, evidence, and avenues for research. *Child Developmental Perspectives*, 7, 2, 126–31.

de Waal, F. 1991. "The chimpanzee's sense of social regularity and its relation to the human sense of justice." *American Behavioral Scientist*, 34, 3, 335–49.

———. 1997. *Good natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, MA: Harvard University Press.

———. 1998. *Chimpanzee Politics. Power and Sex among Apes*. Revised edition. Baltimore, MD: Johns Hopkins University Press.

Dickie, I. 2017. *Fixing Reference*. Oxford: Oxford University Press.

Diesendruck, G., Gelman, S. A., and Lebowitz, K. 1998. "Conceptual and linguistic biases in children's word learning." *Developmental Psychology*, 34, 823–39.

Diesendruck, G., Hall, D., Graham, S., 2006. "Children's use of syntactic and pragmatic knowledge in the interpretation of novel adjectives." *Child Development*, 77, 16–30.

Diesendruck, G., and Markson, L., 2001. "Children's avoidance of lexical overlap: A pragmatic account." *Developmental Psychology*, 37, 630–41.

Donellan, K. 1966. "Reference and definite descriptions." *The Philosophical Review*, 75, 281–304.

———. 1970. "Proper names and identifying descriptions." *Synthese*, 12, 335–58.

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

———, 1993. "What do I know when I know a language." In M. Dummett. *The Seas of Language*, 94–105. Oxford: Clarendon Press.

Doris, J., and Plakias, A. 2008. "How to argue about moral disagreement: Evaluative diversity and moral realism." In *Moral Psychology, Vol. 2: The Cognitive Science of Morality: Intuition and Diversity* edited by W. Sinnott-Armstrong, 303–32, Cambridge, MA: MIT Press.

Dunfield, K. A., Kuhlmeier, V. A., O'Connel, L., Kelley, E., 2011. "Examining the diversity of prosocial behavior: Helping, sharing, and comforting in infancy." *Infancy*, 16, 227–47.

Eimas, P., Siqueland, E., Jusczyk, P., and Vigorito, J., 1971. "Speech perception in infants." *Science*, 171, 303–306.

Eisenberg, N. and Fabes, R. A. 1998. "Prosocial development." In *Social, Emotional, and Personality Development* volume 3 of the *Handbook of Child Psychology* (5th edition) edited by N. Eisenberg, 701–78, New York: Wiley.

Eisenberg, N., Spinrad, T., and Knafo-Noam, A. 2015. "Prosocial development." In *Handbook of Child Psychology and Developmental Science 7th edition* edited by W. Damon, 610–56, Hoboken, NJ: Wiley.

Elbourne, P. 2005. *Situations and Individuals*. Cambridge, MA: MIT Press.

———. P. 2008. "Demonstratives as individual concepts." *Linguistics and Philosophy*, 31, 409–66.

———. 2013. *Definite Descriptions*. Oxford: Oxford University Press.

Elman, J. 2003. "Generalizations from Sparse Input." In *Proceedings of the 38th Annual Meeting of the Chicago Linguistic Society*. Chicago: Chicago University Press.

Embick, D. and Marantz, A. 2008. "Architecture and blocking." *Linguistic Inquiry*, 39, 1, 1–53.

Enoch, D. 2011. *Taking Morality Seriously*. Oxford: Oxford University Press.

Erickson, T. A., and Matteson, M. E. 1981. "From words to meanings: A semantic illusion." *Journal of Verbal Learning and Verbal Behavior*, 20, 540–52.

Etcoff, N. L., and Magee, J. J. 1992. "Categorical perception of facial expressions." *Cognition*, 44, 224–40.

Everett, A. 2000. "Referentialism and empty names." In *Empty Names, Fiction, and the Puzzles of Non-existence* edited by A. Everett and T. Hofweber, 2000, 37–60, CSLI: University of Chicago Press.

Ezcurdia, M., and Stainton, R. (eds.) 2013. *The Semantics-Pragmatics Boundary in Philosophy*. Ontario: Broadview Press.

Ferguson, C., and Farwell, C., 1975. "Words and sounds in early language acquisition." *Language*, 51, 419–39.

Fernald, A. 1993. "Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages." *Child Development*, 64, 657–74.

Ferreira, F. 2003. "The misinterpretation of noncanonical sentences." *Cognitive Psychology*, 47, 164–203.

Ferreira, F., Bailey, K. G. D., and Ferraro, V. 2002. "Good-enough representations in language comprehension." *Current Directions in Psychological Science*, Vol. 11, No. 1, 11–15.

Ferreira, F., Ferraro, V., and Bailey, K. G. D. 2002. "Good enough representations in language comprehension." *Current Directions in Psychological Science*, 11, 11–15.

Ferreira, F., and Patson, N. D. 2007. "The 'good enough' approach to language comprehension." *Language and Linguistics Compass*, 1, 1–2, 71–83.

Fisher, C., Hall, D. G., Raskowitz, S., and Gleitman, L. R., 1994. "When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth." *Lingua*, 92, 333–76.

Fodor, J. A., 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.

———. 2001. "Language, thought and compositionality." *Mind & Language*, 16, 1–15.

———. 2008. *LOT2 The language of Thought Revisited*. Oxford: Clarendon Press.

Fodor, J. A. and Lepore, E. 2002. *The Compositionality Papers*. Oxford: Oxford University Press.

Foucault, M., 1973. *The Order of Things*. Translated by Alan Sheridan. New York: Vintage.

Frege, G. 1879. *Begriffschrift, eine der arithmetischen nachgebildeten Formelsprache des reinen Denkens,* Halle: Nebert.

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

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

Friend, S. 2009. "The great beetle debate: A study in imagining with names." *Philosophical Studies*, 153, 2, 183–211.

Gadamer, G. 1989. *Truth and Method*. 2nd revised edition (1st edition, 1975, translated by W. Glen–Doepel, edited by J. Cumming and G. Barden) revised translation by J. Weinsheimer and D. G. Marshall. New York: Crossroad.

García-Carpintero, M., and Kölbel, M. 2008. *Relative Truth*. Oxford: Oxford University Press.

García-Ramírez, E., and Shatz, M. 2011. "On problems with descriptivism: Psychological assumptions and empirical evidence." *Mind & Language*, 26, 1, 53–77.

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.

Gentner, D., Holyoak, K., and Kokynov, B. (eds.) 2001. *The Analogical Mind*. Cambridge, MA: MIT Press.

Geraci, A., and Surian, L. 2011. "The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources." *Developmental Science*, 14, 5, 1012–20.

Geurts, B. 1998: "The mechanisms of denial." *Language*, 74, 2, 274–307.

Gibson, E., and Pearlmutter, N., 2011. (eds.) *The Processing and Acquisition of Reference*. Cambridge, MA: MIT Press.

Gigerenzer, G. 2008. *Rationality for Mortals: How People Cope with Uncertainty*. Oxford: Oxford University Press.

———. 2008b. "Moral intuition = fast and frugal heuristics?" In *Moral Psychology Vol. 2. The Cognitive Science of Morality: Intuition and Diversity* edited by W. Sinnott-Armstrong, 1–26, Cambridge, MA: MIT Press.

Gigerenzer, G., Todd, P. M., and the ABC Research Group. 1999. *Simple Heuristics that Make Us Smart*. Oxford: Oxford University Press.

Givón, T. 1993. *English Grammar: A Function-Based Introduction*. Amsterdam: Benjamins.

Gleitman, L. 1990. "The structural sources of verb meanings." *Language Acquisition*, 1, 3–55.

Gómez, J. C. 2008. "The evolution of pretence: From intentional availability to intentional non-existence." *Mind & Language*, 23, 5, 586–606.

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.

Greene, J. 2005a. "Emotion and cognition in moral judgment: Evidence from neuroimaging." In *Neurobiology of Human Values. Research and Perspectives in Neurosciences* edited by J. P. Changeux, A. R. Damasio, W. Singer, and Y. Christen, 57–66, Berlin, Heidelberg: Springer.

———. 2005b. "Cognitive neuroscience and the structure of the moral mind." In *The Innate Mind. Structure and Contents* edited by S. Laurence, P. Carruthers, and S. Stich, 338–52, New York: Oxford University Press.

Grice, H. P. 1967. "Logic and conversation." In *Studies in the Way of Words* by H. P. Grice, 1989, 22–40, Cambridge, MA: Harvard University Press.

———. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Haidt, J. 2001. "The emotional dog and its rational tail: A social intuitionist approach to moral judgment." *Psychological Review*, 108, 814–34.

———. 2007. "The new synthesis in moral psychology." *Science*, 316, 998–1002.

Hall, G., 1996. "Preschooler's default assumptions about word meaning: Proper names designate unique individuals." *Developmental Psychology*, 32, 1, 177–86.

Hall, G., 2002. "Semantics and the acquisition of proper names." In *Language, Logic, and Concepts* edited by R. Jackendoff, P. Bloom, and K. Wynn, 337–72. Cambridge, MA: MIT Press.

Hall, G., and Graham, S. 1999. "Lexical form class information guides word-to-object mapping in preschoolers." *Child Development*, 70, 1, 78–91.

Hall, G., Waxman, S. R., Bredart, S., and Nicolay, A. C. 2003. "Preschoolers' use of form class cues to learn descriptive proper names." *Child Development*, 74, 1547–60.

Halle, M., and Marantz, A. 1993. "Distributed morphology and the pieces of inflection." In *The View from Building 20* edited by K. Hale and S. J. Keyser, 111–76, Cambridge, MA: MIT Press.

Halle, M. and Marantz, A. 1994. "Some key features of distributed morphology." In *MIT Working Papers in Linguistics 21: Papers on phonology and morphology* edited by A. Carnie and H. Harley, 275–88, Cambridge, MA.

Halliday, M. A. K. 1975. *Learning How to Mean*. London: Edward Arnold.

Hamlin, J. K. 2013. "Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core." *Current Directions in Psychological Science*, 22, 186–93.

Hamlin, J. K., Ullman, T., Tenenbaum, J. B., Goodman, N., and Baker C. 2013. "The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model." *Developmental Science*, 16, 2, 209–26.

Hamlin, J. K., Wynn, K., and Bloom, P. 2007. "Social evolution by preverbal infants." *Nature*, 450, 557–60.

———. 2010. "Three-month-olds show a negativity bias in their social evaluations." *Developmental Science*, 13, 6, 923–29.

Hamlin, J. K., Wynn, K., Bloom, P., and Mahajan, N. 2011. "How infants and toddlers react to antisocial others." *Proceedings of the National Academy of Sciences*, 108, 19931–36.

Hanna, J. E., Tanenhaus, M. K., Trueswell, J. C. 2003. "The effects of common ground and perspective on domains of referential interpretation." *Journal of Memory and Language*, 49, 43–61.

Hardy, S. A., Walker, L. J., Olsen, J. A., Woodbury, R. D., and Hickman, J. R. 2014. "Moral identity as moral ideal self: Links to adolescent outcomes." *Developmental Psychology*, 50, 1, 45–57.

Hauser, M. 2006. *Moral Minds: How Nature Designed our Universal Sense of Right and Wrong*. New York: Ecco/HarperCollins.

Heibeck, T. H., and Markman, E. M. 1987. "Word learning in children: An examination of fast mapping." *Child Development*, 58, 4, 1021–34.

Heidegger, M. 1962. *Being and Time*. Translated by J. Macquarrie and E. Robinson. Oxford: Basil Blackwell.

Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Dissertation. University of Massachusetts.

Heim, I. and Kratzer, A. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

Henderson, A. M. E., and Woodward, A. L. 2011. "Let's work together: What do infants understand about collaborative goals?" *Cognition*, 121, 12–21.

Hintikka, J., 1983. *The Game of Language. Studies in Game-Theoretical Semantics and its Applications*. Synthese Language Library, Dordrecht: Reidl.

Hirsh-Pasek, K., Treiman, R., and Schneiderman, M. 1984. "Brown & Hanlon revisited: Mothers' sensitivity to ungrammatical forms." *Journal of Child Language*, 11, 1, 81–88.

Hirsh-Pasek, K., Golinkoff, R., and Naigles, L. 1996. "Young children's use of syntactic frames to derive meaning. In *The Origins of Grammar* edited by K. Hirsh-Pasek and R. Golinkoff, 123–58, Cambridge, MA: MIT Press.

Hoff, E. 2006. "How social contexts support and shape language development." *Developmental Review*, 26, 55–88.

Hoff, E., and Shatz, M., eds. 2007. *Blackwell Handbook of Language Development*. Oxford: Blackwell.

Hollich, G. J., Hirsh-Pasek, K., and Golinkoff, R. M. 2000. "Breaking the language barrier: An emergentist coalition model for the origins of word learning." *Monographs of the Society for Research in Child Development*, 65, 3, i–vi, and 1–135.

Horgan, T., and Timmons, M. 2006. *Metaethics after Moore*. Oxford: Oxford University Press.

Humberstone, I. L. 1992. "Direction of fit." *Mind*, 101, 401, 59–83.

Hume, D. 2007. *A Treatise of Human Nature. A Critical Edition*. Edited by D. F. Norton and M. J. Norton, Oxford: Oxford University Press.

Hyams, N. 1986. *Language Acquisition and the Theory of Parameters*. Dordrecht: The Netherlands: Reidel.

Jackendoff, R. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

———, 2007. "A parallel architecture perspective on language processing." *Brain Research*, 1146, 2–22.

Jackson, F. 1998. "Reference and description revisited." *Philosophical Perspectives*, 12, 201–18.

Jacob, P. 2014. "Intentionality." In *The Stanford Encyclopedia of Philosophy* (Winter 2014 edition) edited by Edward N. Zalta, URL=https://plato.stanford.edu/archives/win2014/entries/intentionality/

Jacobs, J. 2002. *Dimensions of Moral Theory: An Introduction to Metaethics and Moral Psychology*. Oxford: Blackwell Publishing.

Jacobson, P. 2014. *Compositional Semantics: An Introduction to the syntax/semantics interface*. Oxford: Oxford University Press.

Janssen, T. M. V. 1997. "Compositionality." In *Handbook of Logic and Language* edited by J. van Benthem and A. ter Meulen, 417–73. Amsterdam: Elsevier.

———. 2001. "Frege, contextuality and compositionality." *Journal of Logic, Language and Information* 10(1): 115–36.

———. 2012. "Compositionality: Its Historic Context." In *The Oxford Handbook of Compositionality* edited by M. Werning, W. Hinzen and E. Machery, 19–45. Oxford: Oxford Univesity Press.

Joyce, R. 2006. *The Evolution of Morality*. Cambridge, MA: MIT Press.

Jusczyk, P. 1997. *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.
Kahane, G. 2011. "Evolutionary debunking arguments." *Noûs*, 45, 1, 103–25.
Kalderon, M. E. 2004. "Open questions and the manifest image." *Philosophy and Phenomenological Research*, 68, 2, 25–89.
Kaminski, J., Call, J., and Fischer, J. 2004. "Word learning in a domestic dog: Evidence for fast mapping." *Science*, 304, 1682–83.
Kamp, H. 1981. "A theory of truth and semantic representation." In *Formal Methods in the Study of Language* edited by I. Groenendijk, T. Janssen, and M. Stokhof, 1–14. Amsterdam: CWI.
Kamp, H. and Reyle, U. 1993. *From Discourse to Logic*. Dordrecht: Kluwer.
Kant, I. 1999. *Critique of Pure Reason*. Translated by P. Guyer and A. Wood. Cambridge: Cambridge University Press.
Kaplan, D. 1989. "Demonstratives. An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals." In *Themes from Kaplan*. J. Almog, H. Wettstein, and J. Perry (eds.), 481–566. Oxford: 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.
Kenward, B., and Dahl, M. 2011. "Preschoolers distribute scarce resources according to the moral valence of recipients' previous actions." *Developmental Psychology*, 47, 4, 1054–64.
Killen, M. and Smetana, J. 2006. *Handbook of Moral Development*. Mahwah, NJ: Erlbaum.
King, J. C. 2017. "W(h)ither Semantics!(?)." *Noûs*, 1–24. Accessed April 24, 2017. doi: 10.1111/nous.12195.
Kitcher, P. 2011. *The Ethical Project*. Cambridge, MA: Harvard University Press.
Knight, B. P., Bernal, M., and Carlo, G. 1995. "Socialization and the development of cooperative, competitive, and individualistic behaviors among Mexican American children." In *Meeting the Challenge of Linguistic and Cultural Divesity in Early Childhood Education, Vol. 6: Yearbook in Early Childhood Education* edited by E. E. Garcia and B. M. McLaughlin, 85–102, New York: Teachers College Press.
Knight, G. P., and Carlo, G. 2012. "Prosocial development among Mexican American youth." *Child Development Perspectives*, 6, 258–63.
Kochanska, G. 2002. "Mutually responsive orientation between mothers and young children: A context for early development of conscience." *Current Directions in Psychological Science*, 11, 191–95.
Kochanska, G., and Kim, S. 2014. "A complex interplay among the parent-child relationship, effortful control, and internalized, rule-compatible conduct in young children: Evidence from two studies." *Developmental Psychology*, 50, 1, 8–21.
Kochanska, G., Koenig, J. L., Barry, R. A., Kim, S., and Yoon, J. E. 2010. "Children's conscience during toddler and preschool years, moral self, and a competent, adaptive developmental trajectory." *Developmental Psychology*, 46, 5, 1320–32.

Kohlberg, L. 1969. "Stage and sequence: The cognitive developmental approach to socialization." In *Handbook of Socialization Theory and Research* edited by D. Goslin, 347–80, Chicago: Rand McNally.

Korta, K. and Perry, J. 2011. *Critical Pragmatics*. Cambridge: Cambridge University Press.

Kovacs, A. M., Teglas, E., and Endress, A. D. 2010. "The social sense: Susceptibility to others' beliefs in human infants and adults." *Science*, 330, 1830–34.

Kratzer, A. 1977. "What *must* and *can* must and can mean." In A. Kratzer. 2012. *Modals and Conditionals*. Oxford: Oxford University Press, 4–20.

———. 1981. "The notional category of modality." In A. Kratzer, 2012. *Modals and Conditionals*. Oxford: Oxford University Press, 27–69.

Kriegel, U. 2007. "Intentional inexistence and phenomenal intentionality." *Philosophical Perspectives, 21, Philosophy of Mind*, 307–40.

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

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

Kuhl, P. K. 2004. "Early Language Acquisition: Cracking the Speech Code." *Nature Reviews Neuroscience*, 5, 831–43.

———. and Miller, J. D. 1975. "Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants." *Science*, 190, 69–72.

———. and Padden, D. M., 1982. "Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques." *Perception and Psychophysics*, 32, 542–50.

Kuhlmeier, V., Wynn, K., and Bloom, P. 2003. "Attribution of dispositional states by 12-month-olds." *Psychological Science*, 14, 5, 402–408.

Kumar, V. 2015. "Moral judgment as a natural kind." *Philosophical Studies*, 172, 2887–910.

Kuperberg, G. R. 2007. "Neural mechanisms of language comprehension: Challenges to syntax." *Brain Research*, 1146, 23–49.

Lagattuta, K. H. 2005. "When you shouldn't do what you want to do: Young children's understanding of desires, rules, and emotions." *Child Development*, 76, 713–33.

Langacker, R. 1987. *Foundations of Cognitive Grammar: Vol.1.* Stanford, CA: Stanford University Press.

Lany, J., and Saffran, J. R. 2010. "From statistics to meaning: Infants' acquisition of lexical categories." *Psychological Science*, 21, 2, 284–91.

Lapsley, D and Carlo, G. 2014. "Moral development at the crossroads: New trends and possible futures [Special issue]." *Developmental Psychology*, 50, 1, 1–7.

Lapsley, D. K., and Hill, P. L. 2009. "The development of the moral personality." In *Personality, Identity, and Character: Explorations in Moral Psychology* edited by D. Narvaez and D. K. Lapsley, 185–213, New York: Cambridge University Press.

Larson, R., and Segal, G. 1995. *Knowledge of Meaning: An Introduction to Semantic Theory*. Cambridge, MA: MIT Press.

Lane, J. D., Wellman, H. M., Olson, S. L., LaBounty, J., and Kerr, D. C. R. 2010. "Theory of mind and emotion understanding predict moral development in early childhood." *British Journal of Developmental Psychology*, 28, 871–89.

Lasersohn, P. 2017. *Subjectivity and Perspective in Truth-Theoretic Semantics*. Oxford: Oxford University Press.

Lassiter, D. 2017. *Graded Modality: Qualitative and Quantitative Perspectives*. Oxford: Oxford University Press.

Leonard, L., Schwartz, R., Morris, B., Chapman, K. 1981. "Factors influencing early lexical acquisition: Lexical orientation and phonological composition." *Child Development*, 52, 882–87.

Leslie, A. 1987. "Pretense and representation: The origins of 'theory of mind.'" *Psychological Review*, 94, 412–26.

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

———, 2002. "Pretense and representation revisited." In *Representation, Memory, and Development: Essays in Honor of Jean Mandler* edited by N. Stein, P. Bauer, and M. Rabinowitz, 103–14. Mahwah, NJ: Lawrence Erlbaum.

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

Lewis, D. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

———. 1970. "General Semantics." In *Philosophical Papers, Vol. 1*, D. Lewis, 1983, 189–229, Oxford University Press.

———. 1973. *Counterfactuals*. Oxford: Blackwell.

———. 1975. "Languages and Language." In *Philosophical Papers, Vol. 1*, D. Lewis, 1983, 163–188, Oxford University Press.

———. 1980. "Index, context, and content." In *Philosophy and Grammar* edited by S. Kanger and S. Öhman, 79–100. Reidel Publishing.

———. 1983. *Philosophical Papers, Vol. 1*. Oxford University Press.

———. 1984. "Putnam's Paradox." *Australasian Journal of Philosophy*, 62, 3, 221–36.

———. 1989. "Dispositional theories of value." Reprinted in *Papers in Ethics and Social Philosophy*, D. K. Lewis, 2000, 68–94, Cambridge: Cambridge University Press.

———. 1992. "Meaning without use: Reply to Hawthorne." *Australasian Journal of Philosophy*, 70, 106–10.

———. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy*, 74, 549–67.

Liberman, A. M., Harris, K. S., Hoffman, H., and Griffith, B. 1957. "The discrimination of speech sounds within and across phoneme boundaries." *Journal of Experimental Psychology*, 54, 358–68.

Lidz, J., and Waxman, S., 2004. "Reaffirming the poverty of stimulus argument: A reply to the replies." *Cognition*, 93, 157–65.

Lidz, J., Waxman, S., and Freedman, J., 2003. "What infants know about syntax but couldn't have learned: Evidence for syntactic structure at 18 months." *Cognition*, 89, B65–B73.

Lillard, A. S. 1993. "Pretend play skills and the child's theory of mind." *Child Development*, 64, 348–71.

Lillard, A. S. 2001. "Pretend play as twin earth: A social-cognitive analysis." *Developmental Review*, 21, 495–531.

Lillehammer, H. 2004. "Moral error theory." *Proceedings of the Australasian Society*, 104, 95–111.

Liu, H.-M., Kuhl, P. K., Tsao, F.-M. 2003. "An association between mothers' speech clarity and infants' speech discrimination skills." *Developmental Science*, 6, F1–F10.

Loeb, D. 1998. "Moral realism and the argument from disagreement." *Philosophical Studies*, 90, 3, 281–303.

Lotze, H. 1874. *Logik*. Leipzig: Hirzel.

Macnamara, J. 1982. *Names for Things: A Study of Human Learning*. Cambridge, MA: MIT Press.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin Books.

Marantz, A. 2010. "Locality domains for contextual allosemy in words." Handout. New York University.

Marantz, A. 2013. "Locality domains for contextual allomorphy across the interfaces." In *Distributed Morphology Today: Morphemes for Morris Halle*, edited by O. Matushansky and A. Marantz, 95–115, Cambridge, MA: MIT Press.

Marcus, G. F. 1993. "Negative evidence in language acquisition." *Cognition*, 46, 53–85.

Marcus, R. B. 1961. "Modalities and intentional languages." *Synthese*, 13, 4, 302–22.

Markman, E. M. 1989. *Categorization and Naming in Young Children*. Cambridge, MA: MIT Press.

———. 1991. "The whole object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meaning." In *Perspectives on Language and Thought: Interrelations in Development* edited by S. Gelman and J. Byrnes, 72–106. Cambridge: Cambridge University Press.

———. 1992. "Constraints on word learning: Speculations about their nature, origins, and domain specificity." In *Modularity and Constraints in Language and Cognition* edited by M. R. Gunnar and M. Maratsos, 59–101. Hillsdale, NJ: Erlbaum.

Markman, E. M., and Jaswal, V. 2004. "Acquiring and using grammatical form class: Lessons from proper-count distinction. In *Weaving a Lexicon* edited by D. G. Hall and S. R. Waxman, 371–410. Cambridge, MA: MIT Press.

Markson, L, and Bloom, P. 1997. "Evidence against a dedicated system for word learning in children." *Nature*, 27, 813–15.

Martin, G. B., and Clark, R. D. 1982. "Distress crying in neonates: Species and peer specificity." *Developmental Psychology*, 18, 3–9.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., and Halwes, T. 1971. "Discrimination in speech and nonspeech modes." *Cognitive Psychology*, 2, 131–57.

Matushansky, O., 2006. "Why Rose is the Rose: On the use of Definite Articles in Proper Names." In O. Bonami and P. Cabredo Hofherr, (eds.). *Empirical Issues in Syntax and Semantics*, *6*, 285–307.

Matushansky, O., 2008. "On the Linguistic Complexity of Proper Names." *Linguistics and Philosophy*, 21, 573–627.

Miller, A. 2014. *Contemporary Metaethics: An Introduction. 2nd edition*. Cambridge: Blackwell Publishing.

Mintz, T. 2002. "Category induction from distributional cues in an artificial language." *Memory and Cognition*, 30, 678–86.

Mintz, T., Newport, E., and Bever, T. 2002. "The distributional structure of grammatical categories in speech to young children." *Cognitive Science*, 26, 393–424.

Montague, R., 1970a. "English as a formal language." Reprinted in R. Montague, 1974. *Formal Philosophy: Selected Papers of Richard Montague* edited by R. Thomasson, 1974, 188–221. New Haven: Yale University Press.

———. 1970b. "Universal grammar." Reprinted in R. Montague, 1974. *Formal Philosophy: Selected Papers of Richard Montague* edited by R. Thomasson, 1974, 222–46. New Haven: Yale University Press.

———. 1970c. "The proper treatment of quantification in ordinary English." Reprinted in R. Montague, 1974. *Formal Philosophy: Selected Papers of Richard Montague* edited by R. Thomasson, 1974, 247–70. New Haven: Yale University Press.

———. 1974. *Formal Philosophy: Selected Papers of Richard Montague*. R. Thomasson (ed.). New Haven: Yale University Press.

Moore, G. E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.

Mueller Gathercole, V. C. and Hoff, E. 2007. "Input and the Acquisition of Language: Three Questions." In *Blackwell Handbook of Language Development* edited by E. Hoff and M. Shatz, 107–27. Oxford: Blackwell.

Nadig, A. S., and Sedivy, J. C., 2002. "Evidence of perspective-taking constraints in children's on-line reference resolution." *Psychological Science*, 13, 4, 329–36.

Naigles, L., and Swensen, L. 2007. "Syntactic supports for word learning." In *Blackwell Handbook of Language Development* edited by E. Hoff and M. Shatz, 212–32. Oxford: Blackwell.

Natsopoulos, D. 1985. "A verbal illusion in two languages." *Journal of Psycholinguistic Research*, 14, 4, 385–97.

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

———. 1974: "Concept, word, and sentence: Interrelations in acquisition and development." *Psychological Review*, 81, 267–85.

Nelson, L. H. 1996. "Empiricism without dogmas." In *Feminism, Science, and the Philosophy of Science* edited by L. H. Nelson and J. Nelson, 1996, 95–120, Dordrecht: Kluwer.

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

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

Nieuwland, M. S., and van Berkum, J. J. A. 2005. "Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension." *Cognitive Brain Research*, 24, 691–701.

Nucci, L, Narvaez, D., and Krettenauer, T. 2014. *Handbook of Moral and Character Education. 2nd Edition*. New York: Routledge.

Nunberg, G., Sag, I., and Wasow, T., 1994. "Idioms." *Language,* 70, 3, 491–538.

Otten, M., and van Berkum, J. J. A. 2007. "What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse dependent N400 effect." *Brain Research*, 1153, 166–77.

Pagin, P. 2005. "Compositionality and context." In *Contextualism in Philosophy: Knowledge, Meaning, and Truth* edited by G. Preyer and G. Peter, 303–48. Oxford: Clarendon Press.

Pagin, P. 2012. "Communication and the complexity of semantics." In *The Oxford Handbook of Compositionality* edited by M. Werning, W. Hinzen, and E. Machery, 510–29. Oxford: Oxford University Press.

Partee, B. H. 1975. "Montague grammar and transformational grammar." *Linguistic Inquiry*, Vol. VI, No. 2, 203–300.

———. 1982. "Believe-sentences and the limits of semantics." In *Processes, Beliefs, and Questions* edited by S. Peters and E. Saarinen, 87–106. Synthese Language Library. Dordrecht: Reidel.

———. 1988. "Semantic facts and psychological facts." *Mind & Language*, 3, 43–52.

———. 1984. "Compositionality." In *Varieties of Formal Semantics*. F., Landman and F. Veltman (eds.), No. 3 GRASS, Dordrecht: Foris, 281–311.

———. 2004. *Compositionality in Formal Semantics*. Oxford: Blackwell Publishing.

Pellizoni, S, Siegal, M., and Surian, L. 2010. "Contact principle and utilitarian moral judgments in young children." *Developmental Science*, 13, 2, 265–70.

Penn, D. C., Holyoak, K. J., and Povinelli, D. J., 2008. "Darwin's Mistake: Explaining the Discontinuity between Human and Non-human 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 & Language*, 22, 5, 475–513.

Perry, J. 1979. "The problem of the essential indexical." *Noûs*, 13, 1, 3–21.

Perry, J. 1993. *The Problem of the Essential Indexical and Other Essays*. Oxford: Oxford University Press.

Perry, J. 2001. *Reference and Reflexivity*, CSLI Publications: University of Chicago Press.

Peterson, C., Slaughter, V., Moore, C., and Wellman, H. M. 2016. "Peer social skills and theory of mind in children with autism, deafness, or typical development." *Developmental Psychology*, 52, 1, 46–57.

Piaget, J. 1932. *The Moral Judgment of the Child*. London: Kegan Paul, Trench, Trübner & Co.

Pierrehumbert, J. B. 2003. "Phonetic diversity, statistical learning and acquisition of phonology." *Language and Speech*, 46, 115–54.

Pinker, S. 1994. *The Language Instinct*. London: Allen Lane.

Platts, M. 1979. *Ways of Meaning*. London: Routledge and Kegan Paul.

Podlaskowski, A. 2018. "Compositionality and the prospect of pluralistic semantic theory." *Australasian Journal of Philosophy*. DOI: 10.1080/00048402.2018.1456554 .

Pollak, S., and Kistler, D., 2002. "Early experience is associated with the development of categorical representations for facial expressions of emotion." *Proceedings of the National Academy of Sciences of the USA*, 99, 9072–76.

Prasada, S. Ferenz, K., and Haskell, T. 2002. "Conceiving of entities as objects and as stuff." *Cognition*, 83, 141–65.

Preyer, G. and Peter, G. 2005. *Contextualism in Philosophy. Knowledge, Meaning, and Truth*. Oxford: Oxford University Press.

Putnam, H. 1975. "The meaning of 'meaning.'" In *Mind, Language, and Reality*. *Philosophical Papers*, Vol. 2, 215–71, Cambridge: Cambridge University Press.

Quine, W. V. O. 1953. *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

———. 1960. *Word and Object*. Cambridge. MA: MIT Press.

Railton, P. 1986. "Moral realism." *The Philosophical Review*, 95, 2, 163–207.

Railton. P. 2014. "The affective dog and its rational tail: Intuition and attunement." *Ethics*, 124, 4, 813–59.

Ramirez, M., Castaneda, A. 1974. *Cultural Democracy, Biocognitive Development, and Education*. New York: Academic Press.

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

———. 2010. *Truth-Conditional Pragmatics*. Oxford: Oxford University Press.

———. 2012. "Compositionality, flexibility and context dependence." In *The Oxford Handbook of Compositionality* edited by M. Werning, W. Hinzen, and E. Machery, 175–91. Oxford: Oxford Univesity Press.

———. 2012b. *Mental Files*. Oxford: Oxford University Press.

Redington, M., Chater, N., and Finch, S. 1998. "Distributional information: A powerful cue for acquiring syntactic categories." *Cognitive Science*, 22, 425–69.

Reimer, M. 2001a. "The problem of empty names." *Australasian Journal of Philosophy*, 79, 4, 491–506.

———. 2001b. "A "Meinongian" solution to a Millian Problem." *American Philosophical Quarterly*, 38, 3, 233–48.

Roberts, C., 1998: "Information structure in discourse: Towards an integrated formal theory of pragmatics." Ohio State University: typescript.

———. 2002. "Demonstratives as definites." In *Information Sharing* edited by Kees van Deemter and Roger Kibble, 1–48. Stanford, CA: CSLI.

———. 2004. "Context in dynamic interpretation." In *The Handbook of Pragmatics* edited by L. R. Horn and G. Ward, 197–220, Oxford: Blackwell Publishing.

Roth-Hanania, R., Davidov, M., and Zahn-Waxler, C. 2011. "Empathy development from 8 to 16 months: Early signs of concern for others." *Infant Behavior & Development*, 34, 3, 447–58.

Russell, B. 1905. "On denoting." *Mind*, 14, 479–93.

Ryle, G. 1946. "Knowing how and knowing that: The presidential address." *Proceedings of the Aristotelian Society*, 46, 1–16.

Sabbagh, M. A., and Baldwin, D. A. 2001. "Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development." *Child Development*, 72, 4, 1054–70.

Saffran, J. 2001a. "Words in a sea of sounds: The output of statistical learning." *Cognition*, 81, 149–69.

———. 2001b. "The use of predictive dependencies in language learning." *Journal of Memory and Language*, 44, 493–515.

———. 2002. "Constraints on statistical language learning." *Journal of Memory and Language*, 47, 172–96.
———. 2003. "Statistical language learning: Mechanisms and constraints." *Current Directions in Psychological Science*, 12, 110–14.
Saffran, J., Aslin, R., and Newport, E. 1996. "Statistical learning by 8-month-old infants." *Science*, 274, 1926–28.
Saffran, J., Senghas, A., and Trueswell, J. C. 2001. "The acquisition of language by children." *Proceedings of the National Academy of Sciences of the United States of America*, 93, 23, 12874–75.
Saffran, J., and Thiessen, E. 2007. "Domain General Learning Capacities." In *Blackwell Handbook of Language Development* edited by E. Hoff and M. Shatz, 68–86. Oxford: Blackwell.
Sainsbury, M. 2005. *Reference without Referents*. Oxford: Oxford University Press.
———. 2010: *Fiction and Fictionalism*. London: Routledge.
Salmon, N. 1986. *Frege's Puzzle*. California: Ridgeview Publishing.
———. 1989. "Illogical belief." *Philosophical Perspectives Vol. 3, Philosophy of Mind and Action Theory*, 243–85.
———. 1998. "Nonexistence." *Nous*, 32:3, 277–319.
Salomo, D., and Liszkowski, U. 2013. "Sociocultural Settings Influence the Emergence of Prelinguistic Deictic Gestures." *Child Development*, 84, 4, 1296–307.
Sampson, G. 2002. "Exploring the richness of the stimulus." *The Linguistic Review*, 19, 73–104.
Sanford, A. J., and Sturt, P. 2002. "Depth of processing in language comprehension: Not noticing the evidence." *Trends in Cognitive Sciences*, 6, 9, 382–86.
Saul, J. M. 1997. "Substitution and simple sentences." 57, 2, 102–8.
Saxton, M., Backley, P., and Gallaway, C. 2005. "Negative input for grammatical errors: Effects after a lag of 12 weeks." *Journal of Child Language*, 32, 643–72.
Schiffer, S. 1987. *Remnants of Meaning*. Cambridge, MA: MIT Press.
Schmidt, M. F. H., and Sommerville, J. A. 2011. "Fairness expectations and altruistic sharing in 15-month-old human infants." *PLoS One*, 6, 10, e23223. doi:10.1371/journal.pone.0023223.
Schroeder, M. 2012. "Philosophy of language for metaethics." In *The Routledge Companion to the Philosophy of Language* edited by G. Russell and D. Graff Fara, 2012, 705–15, New York: Routledge.
Schwartz, R., and Leonard, L., 1982. "Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition." *Journal of Child Language*, 9, 319–36.
Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., and Carlson, G. N. 1999. "Achieving incremental semantic interpretation through contextual representation." *Cognition*, 71, 109–47.
Segal, G. 2001. "Two theories of names." *Mind & Language*, 16, 547–63.
Sellars, W. 1997. *Empiricism and the Philosophy of Mind*. Cambridge, MA: Harvard University Press.
———. 2007. *In the Space of Reasons*. K. Scharp and R. Brandom, (eds.), Cambridge, MA: Harvard University Press.

Semenza, C. 2009. "The Neuropsychology of Proper Names." *Mind & Language*, 24, 4, 347–69.

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.

Shafer-Landau, R. 1994. "Ethical disagreement, ethical objectivism and moral indeterminacy." *Philosophy and Phenomenological Research*, 54, 2, 331–44.

Shatz, M. 1981. "Learning the rules of the game: Four views of the relation between grammar acquisition and social interaction." In *The Child's Construction of Language* edited by W. Deutsch, 17–38. London: Academic Press.

———. 1987. "Bootstrapping Operations in Child Language." In K. E. Nelson and A. van Kleeck, (eds.) *Children's Language, Vol. 6*, 1–22. Hillsdale, NJ: Erlbaum.

———. 1992. "A forward or backward step in the search for an adequate theory of language acquisition." *Social Development*, 1, 151–54.

———. 1994. *A Toddler's Life: Becoming a Person*. Oxford: Oxford University Press.

———. 2007a. "On the development of the field of language development." In *Blackwell Handbook of Language Development* edited by E. Hoff and M. Shatz, 1–15. Oxford: Blackwell.

———. 2007b. "Revisiting *A Toddler's Life* for *The Toddler Years:* Conversational participation as a tool for learning across knowledge domains." In *Socioemotional Development in the Toddler Years: Transitions and Transformations* edited by C. A. Brownell and C. B. Kopp, 241–57, New York: Guilford Press.

———. 2008. "Language as a consequence and enabler of the exercise of higher-order relational capabilities." *Behavioral and Brain Sciences*, 31, 2, 145–46.

Silk, A. 2016. *Discourse Contextualism*. Oxford: Oxford University Press.

Skinner, B. F. 1957. *Verbal Behavior*. New York: Appleton-Century-Crofts.

Smith, M. 1994. *The Moral Problem*. Oxford: Wiley Blackwell.

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

Soja, N., Carey, S., and Spelke, E. S. 1991. "Ontological categories guide young children's inductions of word meaning: Object terms and substance terms." *Cognition*, 38, 431–45.

Soja, N., Carey, S., and Spelke, E. 1992: "Perception, ontology and word meaning." *Cognition*, 45, 101–107.

Spelke, E. S. 1994. "Initial knowledge: Six suggestions." *Cognition*. 50, 431–45.

Sperber, D., and Wilson, D. 1995. *Relevance: Communication and Cognition*. 2nd edition. Oxford: Blackwell.

Stalnaker, R. 1968. "A theory of conditionals." In *Studies in Logical Theory* edited by N. Rescher, 98–112, Oxford: Blackwell.

———. 1978. "Assertion." In *Context and Content: Essays on Intentionality in Speech and Thought*, R. Stalnaker, 1999, 78–95, Oxford: Oxford University Press.

———. 1999. *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

———. 2014. *Context*. Oxford: Oxford University Press.
Stanley, J. 1997. "Names and rigid designation." In *A Companion to the Philosophy of Language* edited by R. Hale and C. Wright, 555–85, Oxford: Blackwell.
Stanley, J. 2007. *Language in Context*. Oxford: Oxford University Press.
Stanley, J., and Szabó, Z., 2000. "On Quantifier Domain Restriction." *Mind & Language*, 15, 219–61.
Stevenson, C. L. 1944. *Ethics and Language*. New Haven: Yale University Press.
———. 1963. *Facts and Values*. New Haven: Yale University Press.
Steward, M., and Steward, D. 1973. "The observation of Anglo-, Mexican-, and Chinese-American mothers teaching their young sons." *Child Development*, 44, 2, 329–37.
Stoel-Gammon, C., and Sosa, A. 2007. "Phonological Development." In *Blackwell Handbook of Language Development* edited by E. Hoff and M. Shatz, 238–56. Oxford: Blackwell.
Street, S. 2006. " A Darwinian dilemma for realist theories of value." *Philosophical Studies*, 127, 1, 109–66.
Sturgeon, N. 1985. "Moral explanations." In *Ethical Theory 1: The Question of Objectivity* edited by J. Rachels, 1998, 180–209, Oxford: Oxford University Press.
Super, C. M. and Harkness, S. 1997. "The cultural structuring of child development." In *Handbook of Cross-Cultural Psychology, Vol. 2: Brain Processes and Human Development* edited by J. W. Berry, P. R. Dasen, and T. S. Saraswathi, 1–39, Boston: Allyn & Bacon.
———. 2002. "Culture structures the environment for development." *Human Development*, 45, 270–74.
Szabó, Z., 2005. *Semantics vs. Pragmatics*. Oxford: Oxford University Press.
———. 2012. "The case for compositionality." In *The Oxford Handbook of Compositionality* edited by M. Werning, W. Hinzen, and E. Machery, 64–80. Oxford: Oxford Univesity Press.
———. 2013. *Problems of Compositionality*. New York: Routledge.
Tabor, W., Galantucci, G., and Richardson, D. 2004. "Effects of merely local syntactic coherence on sentence processing." *Journal of Memory and Language*, 50, 355–70.
Tager-Flusberg, H. 2007. "Atypical language development: Autism and other neurodevelopmental disorders." In *Blackwell Handbook of Language Development* edited by E. Hoff and M. Shatz, 432–53, Oxford: Blackwell.
Tarski, A. 1932. "The concept of truth in formalized languages." In *Logic, Semantics, Metamathematics* edited by J. Corcoran, 1983, 152–278, Indianapolis, IN: Hackett Publishing Company Inc.
Taylor, K. 2000. "Empty names without compromise: A referentialist semantics for empty names." In *Empty Names, Fiction, and the Puzzles of Non-Existence* edited by A. Everett and T. Hofweber, 2000, 17–36, CSLI Publications: University of Chicago Press.
Thomasson, A. L. 2003. "Speaking of fictional characters." *Dialectica*, 57, 2, 207–26.
Tomasello, M. 1992. *First Verbs: A Case Study in Early Grammatical Development*. Cambridge: Cambridge University Press.

———. 2000. "Do young children have adult syntactic competence?" *Cognition*, 74, 209–53.
———. 2003. *Constructing a Language*. Cambridge, MA: Harvard University Press.
———. 2004. "Syntax or semantics? Response to Lidz et.al." *Cognition*, 93, 139–40.
Townsend, D. J., and Bever, T. G. 2001. *Sentence Comprehension the Integration of Habits and Rules*. Cambridge: MA. MIT Press.
Travis, C. 1997. "Pragmatics." In *A Companion to the Philosophy of Language* edited by B. Hale and C. Wright, 87–107, Oxford: Blackwell.
Trendelenburg, A. 1840. *Logische Untersuchungen*. Berlin: Bethge.
Trueswell, J. C., Tanenhaus, M. K., Kello, C., 1993. "Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528–53.
Valentine, T., Brennen, T., and Brédart, S., 1996: *The Cognitive Psychology of Proper Names*. London: Routledge.
Van Berkum, J. J. A. 2008. "Understanding sentences in context: What brain waves can tell us." *Current Directions in Psychological Science*, 17, 6, 376–80.
Van Berkum, J. J. A., Brown, C. M., and Hagoort, P. 1999. "When does gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research*, 28, 555–71.
Van Berkum, J. J. A., Hagoort, P., and Brown C. M. 1999. "Semantic integration in sentences and discourse: Evidence from the N400." *Journal of Cognitive Neuroscience*, 11:6, 657–71.
Van Berkum, J. J. A., Koornneef, A. W., Otten, M., and Nieuwland, M. S. 2007. "Establishing reference in language comprehension: An electrophysiological perspective." *Brain Research*, 1146, 158–71.
Van Berkum, J. J. A., van den Brink, D., Tesink, C., Kos, M., and Hagoort, P. 2008. "The neural integration of speaker and message." *Journal of Cognitive Neuroscience*, 20, 580–191.
Van Herten, M., Kolk, H. H. J., and Chwilla, D. J. 2005. "An ERP study of P600 effects elicited by semantic anomalies." *Cognitive Brain Research*, 22, 241–55.
Van Valin, R. 1993. "A synopsis of role and reference grammar." In *Advances in Role and Reference Grammar* edited by R. Van Valin, 1-164. Amsterdam: Benjamins.
Velleman, D. J. 1992. "The guise of the good." *Nous*, 26, 1, 3–26.
Walton, K. 1990. *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.
———. 2000. "Existence as metaphor?" In *Empty Names, Fiction, and the Puzzles of Non-Existence* edited by A. Everett and T. Hofweber, 2000, 69–94, CSLI Publications: University of Chicago Press.
———. 2003. "Restricted quantification, negative existentials, and fiction." *Dialectica*, 57, 2, 239–42.
Wang, X., Kadia, S. 2001. "Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat." *Journal of Neurophisiology*, 86, 2616–20.

Wason, P., and Reich, S. S. 1979. "A verbal illusion." *Quarterly Journal of Experimental Psychology*, 31, 591–97.

Waxman, S. R. and Booth, A. E. 2000. "Principles that are invoked in the acquisition of words, but not facts." *Cognition*, 77, B33–B43.

Waxman, S. R. 2004. "Everything had a name, and each name gave birth to a new thought: Links between early word learning and conceptual organization." In *Weaving a Lexicon* edited by D. G. Hall and S. R. Waxman, 295–335. Cambridge, MA: MIT Press.

Wedgwood, R. 2007. *The Nature of Normativity*. Oxford: Oxford University Press.

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

Werheid, K., and Clare, L. 2007. "Are faces special in Alzheimer's disease? Cognitive conceptualization, neural correlates, and diagnostic relevance of impaired memory for faces and names." *Cortex*, 43, 898–906.

Werning, M., Hinzen, W., and Machery, E. (eds.) 2012. *The Oxford Handbook of Compositionality*. Oxford: Oxford University Press.

Westerståhl, D., 2002. "Idioms and Compositionality." In D. Barker-Plummer, D. Beaver, J. van Bentham, and P. Scotto di Luzio, (eds.). 2002. *Words, Proofs, and Diagrams*. Stanford: CSLI Publications. 241–71.

———, 2012. "Compositionality in Kaplan-style Semantics." In *The Oxford Handbook of Compositionality* edited by M. Werning, W. Hinzen, and E. Machery, 192–219. Oxford: Oxford Univesity Press.

Whiting, B. B., and Edwards, C. P. 1988. *Children of Different Worlds: The Formation of Social Behavior*. Cambridge, MA: Harvard University Press.

Whitting, B. B., and Whitting, J. W. M. 1975. *Children of Six Cultures. A Psycho-Cultural Analysis*. Cambridge, MA: Harvard University Press.

Williams, F. L. 1991. "Interfamily economic exchange: A function of culture and economics." *Lifestyles, Family and Economic Issues*, 12, 3, 235–52.

Wilson, D., and Sperber, D. 2002. "Truthfulness and relevance." *Mind*, 111, 443, 583–632.

Wittgenstein, L. 1967. *Philosophical Investigations*. G. E. M. Anscombe, R. Rhees, and G. H. von Wright (eds.), Oxford: Blackwell.

Wood, J. and Marantz, A. 2017. "The interpretation of external arguments." In *The Verbal Domain* edited by R. D'Alessandro, I. Franco, and A. J. Gallego, 255–78, Oxford: Oxford University Press.

Wundt, W. 1880. *Logik. Eine Untersuchung der Principien der Erkenntnis und der Methoden wissenchafliche Forschung*. *Vol. 1. Erkentnisslehre*, 2 vols. Stuttgart: Ferdinand Enke.

Wynn, K. 2008. "Some innate foundations of social and moral cognition." In *The Innate Mind: Vol. 3. Foundations and the Future* edited by P. Carruthers, S. Laurence, and S. Stich, 330–47, Oxford: Oxford University Press.

Yablo, S., 2000: "A paradox of existence." In A. Everett and T. Hofweber, (eds.), 2000: *Empty Names, Fiction, and the Puzzles of Nonexistence* CSLI Publications.

Zadrozny, W. 1994. “From compositional to systematic semantics.” *Linguistics and Philosophy*, 17, 329–42.

Zimmermann, T. 2012. “Compositionality problems and how to solve them.” In *The Oxford Handbook of Compositionality* edited by M. Werning, W. Hinzen, and E. Machery, 81–106. Oxford: Oxford University Press.

Index

About the Author

Eduardo García-Ramírez (PhD, University of Michigan, Ann Arbor) is a Research Fellow at the Institute for Philosophical Research of the National Autonomous University of Mexico (UNAM). His work includes multiple articles showing how empirical studies from cognitive psychology and psycholinguistics can illuminate, and sometimes dissolve, old philosophical problems. He has also published work on metaphysics and metaethics.