FACULDADE DE CIÊNCIAS
SOCIAIS E HUMANAS
UNIVERSIDADE NOVA DE LISBOA

The Rationalization of Utterances

Pedro Rui do Espírito Santo Abreu

Tese de Doutoramento em Filosofia

Data (Julho, 2017)

[**DECLARAÇÕES**]

Declaro que esta tese é o resultado da minha investigação pessoal e independente. O seu conteúdo é original e todas as fontes consultadas estão devidamente mencionadas no texto, nas notas e na bibliografia.

O candidato,

Lisboa, 6 de Julho de 2017

Declaro que esta tese se encontra em condições de ser apreciada pelo júri a designar.

O orientador,

Lisboa, 6 de Julho de 2017

Tese apresentada para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Filosofia, realizada sob a orientação científica do Professor João Sàágua

Apoio financeiro da FCT e do FSE no âmbito do III Quadro Comunitário de Apoio.

ACKNOWLEDGMENTS

First, I would like to thank João Sàágua for all his support and inspiration throughout these years. It started some time ago when he nurtured my nascent interest in philosophy of language and first introduced me to the author that has been central to my work since then, Donald Davidson. He naturally came to be my PhD supervisor, and provided me with invaluable help, direction and encouragement at the various stages of the whole process. He was the dependable and knowledgeable guide that showed me the way around a field and tradition that can be very exotic and challenging to a newcomer. His steady patience and confidence in the progress of my work often contrasted with my own state of mind and pushed me forward at critical times.

I would also like to express my deep gratitude to Ernest Lepore for his hospitality and care during my months in Rutgers, for the perceptive pieces of advice that more than once diverted the course of this investigation, and for his generosity in so carefully reading and commenting the long pieces I sent him. His general approval of the final result means a lot to me. Even where we disagree, it should be evident that he cleared much of the ground upon which I was able to determine and establish my own positions.

I am also very grateful to a small group of people who contributed to this work in one way or another: Paulo Barcelos, Michael Baumtrog, Klaus Gärtner, Ana Falcato, João Fonseca, Marcin Lewiński, Franck Lihoreau, Fabrizio Macagno, Dima Mohammed, Nuno Mora, Bjørn Ramberg, Erich Rast, Olivia Sultanescu and Claudine Verheggen. Sometimes a small and opportune word made all the difference.

Lastly, I thank Susana and Francisco - who has never known me without the peculiar cloud that hangs over PhD candidates' heads - for all their immense love and patience. It has been a very long ride, one of us was born along the way, and you have made it pleasant, serious and worthwhile.

The Rationalization of Utterances

Pedro Rui do Espírito Santo Abreu

RESUMO

Nesta dissertação defendo a tese de que a interpretação linguística requer racionalização, e a concepção geral de linguagem, significado e comunicação que a suporta.

Por “interpretação linguística” entendo o processo através do qual os falantes chegam a atribuições de significados às elocuções dos seus interlocutores. A noção de significado relevante é a de *significado que o falante pretende imprimir à expressão usada*. Por “racionalização” entendo a investigação e atribuição de estados mentais a um agente, que se presume racional, com vista a *fazer sentido* dos estados e acções desse agente. Apela-se a uma noção de racionalidade muito rica e ampla. O *fazer sentido* dos pensamentos e acções de um agente pode ser articulado de várias formas: compreender, explicar, encontrar razões para, acompanhar, prever, e mais. Para ser explícito a respeito da linhagem Davidsoniana desta dissertação, noto que o que está aqui em causa é a ideia de que a interpretação deve ser moldada e guiada pelo Princípio de Caridade. Defendo que a racionalização deve estar presente - mesmo que em diferentes formas e intensidades - em toda a interpretação linguística que genuinamente vise o entendimento entre interlocutores. Isto é assim em todos os tipos de situação comunicativa e em todo os estágios de competência linguística, da interpretação radical e do aprender das palavras da primeira língua, à conversa banal entre dois falantes maduros da mesma língua oficial.

A abordagem racionalizante, e a concepção geral de significado e linguagem que a fundamenta, são contrastadas com duas visões alternativas. Em primeiro lugar temos o *Naturalismo* de Quine com o projecto de compreender as capacidades e práticas linguísticas como um fenómeno natural que deve ser estudado e entendido através dos métodos das ciências naturais. Este projecto integra também algumas propostas a respeito do que possa ser descrito como um método apropriadamente *naturalístico* de interpretação linguística. A segunda alternativa, o *Convencionalismo* de Lepore e Stone, enfatiza o elemento convencional na linguagem e comunicação, e sustenta que, em casos normais, a interpretação linguística é exclusivamente baseada no conhecimento comum do código relevante, dispensando assim o intérprete de racionalização. Eu procuro mostrar que, não obstante méritos vários, nenhuma destas duas alternativas consegue realmente estabelecer alguma tese que comprometa, ou sequer significativamente diminua, a pertinência e valor teórico da perspectiva Davidsoniana aqui adoptada. Em particular, nenhuma delas é bem sucedida na demonstração de que a interpretação linguística pode dispensar racionalização.

ABSTRACT

In this dissertation I defend the claim that linguistic interpretation requires rationalization, and the general conception of language, meaning and communication that supports it.

By "linguistic interpretation" I mean the process through which speakers arrive at meaning ascriptions for their interlocutors' utterances. The relevant notion of meaning is that of the *meaning intended by the speaker for her expression to carry*. By "rationalization" I mean the investigation and ascription of mental states to an agent that aims at *making sense* of the agent's states and actions, which is guided by the presumption that the agent is rational. The notion of *rationality* that I appeal to is a very broad and rich one. The *making sense* of an agent's states and actions may be articulated in various, more or less overlapping, forms: understanding, explaining, finding reasons for, keeping track of, predicting, and more. Being explicit about the Davidsonian lineage of this dissertation, I should note that what is at stake here is the idea that interpretation must be shaped and directed by the *Principle of Charity*. I maintain that rationalization must be present - even if in different forms and intensities - in all linguistic interpretation that genuinely aims at understanding among interlocutors. This is so across all types of communicative situations and stages of linguistic competence, from radical interpretation and the learning of the words of the first language to regular conversations among mature speakers of the same official language.

The rationalizing approach, and the general conception of meaning and language behind it, are contrasted with two alternative views. The first is Quine's *Naturalism* and the project of understanding linguistic abilities and practices as a natural phenomenon to be studied and accounted for using the methods of natural sciences. It also integrates some proposals about what we can describe as appropriately naturalistic methods of linguistic interpretation. The second alternative, Lepore and Stone's *Conventionalism*, emphasizes the conventional element in language and communication, and maintains that regular instances of linguistic interpretation are exclusively based on common knowledge of the relevant code, thus dispensing with rationalization. I argue that, for all that is laudable and insightful about these two alternatives, they do not actually succeed in holding their ground in anything that compromises or significantly diminishes the pertinence and theoretical value of the broadly Davidsonian view that I am here endorsing. In particular, they fail to prove that linguistic interpretation can actually dispense with rationalization.

PALAVRAS-CHAVE: Interpretação, Racionalização, Significado, Linguagem, Davidson

KEYWORDS: Interpretation, Rationalization, Meaning, Language, Davidson

CONTENTS

GENERAL INTRODUCTION 1

Part I
Rationalization

INTRODUCTION TO PART I 9

1. DAVIDSON'S CONCEPTION OF MEANING AND INTERPRETATION 11

1.1. TRUTH-THEORETIC SEMANTICS 11

1.2. RADICAL INTERPRETATION 16

1.2.1. THE LINKS BETWEEN LANGUAGE AND THE WORLD 16

1.2.2. THE ACCOMPANYING CIRCUMSTANCES 17

1.2.3. FROM OBSERVATIONS TO A MEANING THEORY 20

1.2.4. THE PRINCIPLE OF CHARITY 22

1.2.4.1 THE PRINCIPLE OF CORRESPONDENCE 24

1.2.4.2. THE PRINCIPLE OF COHERENCE 31

1.2.5. A GENERAL CONCEPTION OF MEANING AND ACTUAL INTERPRETATION 36

2. JUSTIFICATION OF THE PRINCIPLE OF CHARITY 43

2.1. ARGUING FOR THE PRINCIPLE OF COHERENCE 44

2.1.1. THE ARGUMENT FROM THE NATURE OF THOUGHT 45

2.1.2. THE ARGUMENT FROM SEMANTIC HOLISM 48

2.1.3. THE ARGUMENT FROM INTERPRETATION HOLISM 50

2.1.4 HUMAN IRRATIONALITY 53

2.2. ARGUING FOR THE PRINCIPLE OF CORRESPONDENCE 56

2.2.1. THE ARGUMENT FROM EXORBITANCE AND INCONSISTENCY 61

2.2.2. THE ARGUMENT FROM INTERPRETATION 62

2.2.3. THE ARGUMENT FROM TRIANGULATION 63
2.3 CONCLUSION 65

3. TWO CHALLENGES: FIRST-PERSON KNOWLEDGE AND UNDERDETERMINATION 67
3.1. INTRODUCING THE CHALLENGES 68
3.2. THE CHALLENGE FROM FIRST-PERSON KNOWLEDGE 69
3.2.1. THEORETICITY, ANTI-CARTESIANISM AND INDETERMINACY 70
3.2.2. INTERPRETIVISM 72
3.2.3. THE FIRST CHALLENGE TARGETS INTERPRETIVISM 74
3.2.4. RESISTING THE FIRST CHALLENGE 75
3.3. THE SECOND CHALLENGE 80
3.3.1. MORE THAN TRUTH, CHARITABLE INTERPRETATION 82
3.3.2. DAVIDSON ON UNDERDETERMINATION 83
3.3.3. LEPORE AND LUDWIG ON UNDERDETERMINATION 86
3.3.3.1. RESISTING LEPORE AND LUDWIG'S CASE FOR UNDERDETERMINATION 91
3.4. DIFFERENT CONCEPTIONS OF RADICAL INTERPRETATION 93
3.5. CONCLUSION 99

4. RATIONALIZATION AND LANGUAGE ACQUISITION 101
4.1. RATIONALITY AND RATIONALIZATION 102
4.1.1. RATIONALITY IN WORD LEARNING 104
4.2. EMPIRICAL ACCOUNTS OF WORD LEARNING 109
4.2.1. ASSOCIATIVE PROCESSES IN WORD LEARNING 111
4.2.2. LEXICAL CONSTRAINTS OR PRINCIPLES IN WORD LEARNING 115
4.2.3. MINDREADING AND RATIONALIZATION IN WORD LEARNING 116
4.3. CONCLUSION 124

Part II
Alternatives To Rationalization

INTRODUCTION TO PART II 127

Part II.a.
Rationalization and Naturalization

INTRODUCTION TO PART II.A. 131

5. NATURALISTIC REJECTION OF RATIONALIZATION 133

5.1. EPISTEMOLOGY AND NATURALISM 134
5.1.1. THE ABANDONMENT OF TRADITIONAL EPISTEMOLOGY 135
5.1.2. NATURALIZED EPISTEMOLOGY 136
5.2. CONTROVERSIES WITH QUINE'S NATURALISM 137
5.2.1. QUINE'S CONCEPTION OF SCIENCE 138
5.2.1.1. QUINE'S REJECTION OF MENTAL EXPLANATIONS 141
5.2.1.1.1. AMBIVALENCE AND TENSION 146
5.2.1.1.2. LAST ARGUMENT AGAINST MENTALISM 150
5.3. THE NORMATIVITY OBJECTION 152
5.3.1. QUINE'S RESPONSE 153
5.3.2. NORMATIVE EPISTEMOLOGY IN INTERPRETATION 155
5.4. QUINE'S *COMMON INTERPRETER* 159
5.5. CONCLUSION 160

6 A NATURALISTIC ALTERNATIVE TO RATIONALIZATION 163

6.1. NATURALISTIC ACCOUNT OF OBSERVATION SENTENCES 163
6.1.1. PERCEPTUAL SIMILARITY 166
6.1.2. SIGNALING OBSERVATION SENTENCES AND PERCEPTUAL SIMILARITY 169
6.1.2.1. SOCIAL CRITERION A 169
6.1.2.2. SOCIAL CRITERION B 171
6.1.2.3. INDIVIDUAL CRITERION A 173
6.1.2.4. INDIVIDUAL CRITERION B 174
6.1.2.5. NO SENTENCE SATISFIES THE CRITERIA 175

6.1.2.5.1. COMPLEX OBSERVATIONALITY 180
6.1.3. EXPLAINING PERCEPTUAL SIMILARITY 184
6.1.4. THE FAILURE OF QUINE'S NATURALISTIC ACCOUNT 187
6.2. TRANSLATION NATURALIZED 189
6.2.1. JUST BE NATURAL 189
6.2.2. THE HYBRID METHOD 190
6.2.2.1. NO ROOM FOR THE PRINCIPLE OF CHARITY IN THE TRANSLATION OF OBSERVATION SENTENCES 191
6.2.2.2. EXPLANATORY LACUNAS DO NOT DEFEAT THE HYBRID METHOD 194
6.2.2.3. THE UNDERLYING RATIONALIZATION 196
6.2.3. TRANSLATION THROUGH EMPATHY 198
6.2.4. RATIONALIZATION PREVAILS 200

Part II.b.
Rationalization and Convention

INTRODUCTION TO PART II.B. 205

7. RADICAL ANTI-CONVENTIONALISM: DAVIDSON ON MEANING CONVENTIONS 207

7.1. LEWIS' NOTION OF CONVENTION 208
7.2. FIRST MEANING 214
7.2.1. FIRST MEANING, UTTERER'S MEANING, UTTERANCE-TYPE MEANING AND WHAT IS SAID 219
7.3. DAVIDSON'S ANTI-CONVENTIONALISM 221
7.3.1. ARGUING AGAINST CONVENTIONALISM 224
7.3.2. ARGUING AGAINST SUFFICIENCY 225
7.3.3. ARGUING AGAINST NECESSITY 228
7.3.3.1. THE ARGUMENT FROM RADICAL INTERPRETABILITY 230
7.3.3.1.1. LEPORE AND LUDWIG'S *IN PRINCIPLE* GENERAL DISPENSABILITY 234
7.3.3.2. THE ARGUMENT FROM THE ESSENCE OF COMMUNICATION 236
7.4. CONCLUSION 238

8. RADICAL CONVENTIONALISM: LEPORE AND STONE ON MEANING CONVENTIONS 241

8.1. LEPORE AND STONE'S *IMAGINATION AND CONVENTION* 243
8.1.1. DIRECT INTENTIONALISM VS. PROSPECTIVE INTENTIONALISM 245
8.1.1.1. THREE TYPES OF INTENTIONS 247
8.1.1.2. MEANING DETERMINING INTENTIONS 249
8.1.1.2.1. CONVENTIONS 251
8.1.1.2.2. EXTERNALISM'S INSTITUTIONS 253
8.1.1.2.3. THE CONVERSATIONAL RECORD 254
8.2. MALAPROPISMS AND FAULTY UTTERANCES 257
8.2.1. CHALLENGING STANDARD MEANINGS 260
8.2.1.1. CONVENTIONS AND FAULTY UTTERANCES 260
8.2.1.1.1. UNSUCCESSFUL REJOINDERS 261
8.2.1.2. EXTERNALISM'S INSTITUTIONS AND FAULTY UTTERANCES 265
8.2.1.2.1. KRIPKE 266
8.2.1.2.2. PUTNAM 267
8.2.1.2.3. BURGE 269
8.2.1.2.4. EXTERNALISM ON SEMANTIC AND PSYCHOLOGICAL INTEGRATION 272
8.2.1.2.5. EXTERNALISM AND CHARITY 279
8.2.1.2.6. CONCLUSION ON STANDARD MEANINGS 284
8.2.2. DEFENDING INTENDED MEANINGS 286
8.2.2.1. THE POINT OF OUR SOCIAL USE OF LANGUAGE 291
8.2.2.2. COOPERATION 294
8.2.2.3. FOUNDATION AND INNOVATION 295
8.2.2.4. LANGUAGE ACQUISITION 303
8.2.2.5. CONCLUSION ON INTENDED MEANINGS 306
8.3. CONCLUSION 307

GENERAL CONCLUSION 313

REFERENCES 323

General Introduction

In this dissertation I defend the claim that linguistic interpretation requires rationalization.

By "linguistic interpretation" I mean the process through which agents arrive at meaning ascriptions for their interlocutors' utterances. The relevant notion of meaning is that of the meaning the speaker intends her[1] expression to carry. In the view I will be developing here, this is the most basic notion of linguistic meaning, and one that is crucial for interpretation and communication. The intended meaning will often coincide with the conventional or standard meaning of the expression in the community of reference, but it does not have to.

I am especially interested in the simple cases where the intended meaning coincides with what the speaker ultimately wants to say. I am not taking this investigation into the additional layers of meaning, orthodoxically associated with pragmatics, that are frequently built on top of that elementary form of literal meaning. I am focusing only on this most basic level.

This notion of intended, potentially non-conventional and non-standard, meaning of some expression is none other than Davidson's notion of *first meaning.* A detailed exploration of some of the sensible questions involving this notion and the options and positions just made explicit will be postponed to later in the dissertation.

By "rationalization" I mean the investigation and ascription of mental states to an agent that aims to make sense of the agent's states and actions, and is guided by the presumption that the agent is a rational being. The making sense of an agent's states and actions may be articulated in various, more or less overlapping, forms: understanding, explaining, finding reasons for, keeping track of, predicting, and more.

[1] To reduce ambiguity, I will follow this rule throughout the whole dissertation: female pronouns for *speakers*, male pronouns for *interpreters.*

Here is an interesting illustration of the exercise involving a popular puzzle. I heard it first under a different guise, but it is customarily presented and referred to as *the muddy children puzzle*[2]. Here is one of the simplest versions:

> *Two kids, A and B, who were playing outside are called in by their mother. The mother tells them: "At least one of you has mud on his face, can each of you tell me whether you're muddy or not?" In fact they both have muddy faces, the mother can see it all, naturally, but not the kids. Each kid can only see the other's face, not his own. How can each solve the riddle and deliver the right answer?*

First, let me make it clear that it is not the puzzle in particular, nor its content or theme, that matters to this dissertation, it is rather the method required for its solution. Now, imagine, for instance you are kid A - it does not make a difference since both situations are exactly parallel. Here is how you must reason:

> *I can see that B is muddy. That would be enough to make it true that at least one of us is muddy and so I cannot tell whether I'm muddy myself or not.*
>
> *If B were not muddy, things would be different. In such a case I could infer that I was the muddy one.*

then, after a couple of seconds, ...

> *I can see that B himself doesn't seem to know his answer, he's looking at me just as hesitantly ...*
>
> *oh, but then ... I got it!!*
>
> *... his puzzlement must mean that I'm muddy. If I were not muddy, he could've promptly inferred that he is the muddy one. Since he didn't, it has to be that I'm also muddy.*

As I have said, what matters to me here is the method required for dealing with the question. On the one hand, as with every puzzle in general, it requires a *rational* approach. That is, thinkers must mobilize their theoretical rationality to solve the puzzle. In the end, when they get it, they experience the rational cogency of the right answer, and they experience it very vividly - that is the effect one expects from this kind of puzzle.

---

[2] See, for instance, Ditmarsch and Kooi, 2015, Chapter 3.

On the other hand, less generally - but still not specific to this puzzle - it requires rationalization. That is, to solve the puzzle one has to investigate the mental states of other people, based on incomplete information of various sorts, guided by the assumption that the person is broadly rational in both practical and theoretical respects. As we saw, kid A had to inquire into kid B's thoughts and reasonings starting from the small clues available in the situation together with the assumption that B was rational enough, both theoretically and practically, in his thoughts and action: he had to either assume or figure out that B wanted to give the right answer; that he would have answered before if he could; that he did not answer first because he did not know the answer then; that he believed that A was muddy but was uncertain whether he himself was muddy or not; and so on.

These same two ingredients, rationality and rationalization, are always necessary for linguistic interpretation - although not always to the same degree, not always as explicitly. This is the thesis I am defending in this dissertation. My main focus, however, will be on rationalization. That rationality is present in the exercise I will mostly simply assume. It is the presence of the second ingredient - rationalization - that I will be arguing for.

Rationalization does not need to take such a complex and explicit form as in the above example. There is rationalization involved in much simpler instances of intelligent interaction among rational agents. There is rationalization, for instance, when the interpreter assumes that his interlocutor is aware of the presence of some observable and salient object, feature, or event. There is rationalization when the interpreter follows the interlocutor's pointing gesture in the right direction to discover the object that she intended to call his attention to. There is rationalization when the interpreter expects the speaker to continue a sequence in some particular way. There is rationalization when the interpreter assumes that the sounds coming out of his interlocutor's mouth are intended to be significant and for him to interpret. All of this is natural and obvious to the interpreter, but none of it would be were it not for his promptness in taking his interlocutor to share his rational nature, in a very broad and rich sense of rationality.

Being explicit once more about the Davidsonian spirit of the enterprise, I would like to note that to claim that interpretation requires rationalization is to claim that linguistic interpretation should be addressed only in the context of a general interpretation of the speaker as a rational agent, ascribing thoughts to agents along with meanings, and that the

presumption that the speaker is a rational agent corresponds, of course, to Davidson's Principle of Charity.

When I affirm that interpretation *requires* rationalization, I mean that rationalization must always be present in linguistic interpretation that aims at genuine understanding among interlocutors. Naturally, I am willing to acknowledge that the exercise takes different forms and intensities in different types of situation.

The need for rationalization will be most evident in those situations of *Radical Interpretation,* as Davidson dubbed them, where an interpreter, without the help of dictionaries or bilingual intermediaries, is taking his first steps into the language of his interlocutor. These are of course extremely rare and extraordinary circumstances, but something interestingly close to it is enacted by each child acquiring her first language - all the crucial differences between the two types of situation notwithstanding. Children, I will report in some detail, require an intense use of rationalization when learning the words of their first language.

This initial and more flagrant form of rationalization evolves into something no less pervasive in later stages of proficiency. I argue that even among competent speakers of the same language, even when languages are, to a significant extent, already learned and shared, rationalization will still be needed and will still be present. To be able to secure linguistic understanding, the interpreter will often be required to deviate from the standard or expected course of interpretation. The permanent possibility of such deviations requires a permanent vigilance on the interpreter's part, one that takes the form of rationalization. Furthermore, the occasional figuring out of such deviations will ordinarily require the interpreter to make yet further use of those same abilities, methods and processes.

Instances of rationalization across this wide range of cases will vary in various respects. In particular, as we will see, rationalization may be more or less explicit and conscious, and may be more or less intricate, ingenious and demanding on the interpreter.

What I attempt to achieve in this dissertation is not the complete and decisive demonstration of the truth of a proposition standing by itself, fully meaningful and evaluable no matter the frame of reference. The central claim in this dissertation is integral to a general view of meaning, linguistic practices, and understanding among rational agents. It is not to be assessed, approved or criticized on its own. General views are hardly suited for conclusive

verification of falsification. Both the Davidsonian proposal and the alternatives to be considered are immense and evolving works, grounded on traditions and intuitions of their own, still open to refinement, rectification and addition. Besides, in such contexts, questions of truth and falsity are often entangled and conditional on a number of *theoretical options* (as we might put it): what aspects of the phenomenon we are most interested in, how our notions are delimited, what links with neighboring questions and inquiries we are privileging, and others.

Accordingly, the aim of this dissertation is not to provide conclusive proof of the claim and program endorsed, together with the thorough and implacable rejection of the alternatives. Something more modest and nuanced is in order. My goal here is simply to continue advancing the Davidsonian view of meaning and language in certain particular directions where I found I had something new to offer. This includes proposing and defending particular interpretations of Davidson's claims and positions at various moments of relative obscurity and doubt, responding to standing challenges and obstacles, and establishing new and fertile connections with other inquiries and disciplines. In response to the alternative conceptions considered here, the plan is to identify and stress the specific explanatory potential of the Davidsonian view, and how that value is lost, and with it aspects of the reality under study, to the conceptions that fail to acknowledge that crucial dimension of the linguistic phenomenon central to this dissertation: that interpretation requires rationalization.

This dissertation is composed of eight chapters distributed in two parts. In Part I, Chapters 1-4, I present and argue for the Davidsonian conception of meaning and linguistic communication that supports - and is, in some sense, synthesized in - the thesis being defended here. In Chapter 1, I introduce the basics of Davidson's program on meaning and language. After a brief exposition of Davidson's seminal ideas on truth-theoretic meaning theories, I come to Davidson's conceptual experiment of Radical Interpretation and the Principle of Charity. I emphasize the interdependencies between meaning and thought, and between linguistic interpretation and general interpretation of rational agents. Chapter 2 is devoted to the justification and defense of the Principle of Charity. Chapter 3 addresses two pressing challenges to the Davidsonian conception of meaning and language. Chapter 4 seeks to vindicate some of Davidson's insights by testing his views against recent findings and conclusions in the scientific study of language acquisition and word learning.

In Part II, Chapters 5-8, I examine two alternative approaches to the same questions. In Part II.a I consider Quine's naturalistic project and how it translates into the study of meaning and language. Chapter 5 introduces the general questions and claims of Quine's naturalism, and the consequent criticism of various elements and practices that play an essential role in the rationalization of utterances: traditional epistemology, appeals to rationality, mentalistic idioms. Chapter 6 presents and criticizes Quine's sketches of a naturalistic account of meaning and language use, and of a naturalistic method of interpretation.

In Part II.b, I reflect upon the role of conventions in linguistic interpretation, and whether or not, and to what extent, they can mitigate the need to involve rationalization in the process. I adopt a middle path between Radical Conventionalism and Radical Anti-Conventionalism. In Chapter 7, I resist Davidson's arguments against the necessity of conventions for linguistic communication. In Chapter 8, I resist Lepore and Stone's case in favor of the sufficiency of conventions for linguistic communication.

This is a very brief map of what follows. Each part is preceded by a more substantial introduction to the chapters included in it.

Part I

Meaning and Rationalization

Introduction to Part 1

Interpretation requires rationalization. The exercise of ascribing meanings to speakers must be integrated with that of ascribing propositional states to thinkers, and the whole process must be guided by, or in conformity with, the Principle of Charity. In other words, meaning ascription is a holistic affair, which cannot fail to be accompanied by a more or less explicit, more or less deep, investigation into the speaker's mind, and the whole enterprise is directed and shaped by the interpreter's projection onto the speaker of his own standards of rationality - here understood in a very rich and broad, if vague, way. This is the thesis I am defending here.

This thesis is implied by a certain general and foundational, or metaphysical, conception of meaning that I also find plausible and tentatively adopt, one that is condensed in the claim that *meaning is determined by interpretability*, i.e. that the meaning of any utterance is what a vulgar interpreter in good enough conditions - essentially those delineated by Davidson for his radical interpreter - would interpret it to mean.

In this Part I, I present the background that sustains, motivates and elucidates my thesis, and initiate its defense. In Chapter 1, I introduce Davidson's Radical Interpretation, the conceptual experiment central to all my reflection in these pages, and I start to reveal how it offers favorable ground for insights and conclusions that extend way beyond the farfetched scenario and situation initially depicted by Davidson. In the next three chapters, I endeavor to defend some of these ideas.

Radical Interpretation involves the intense employment of the Principle of Charity. As we will see, the principle compels the interpreter to assume, from the beginning, a fair amount of familiarity with other speakers' minds, including a common rationality and a similar experience of the world in its basic features. The principle plays an indispensable role in the experiment in allowing the interpreter to bridge the gap between observation of the speaker's relation to her environment and the identification of her meanings and thoughts. In Chapter 2, I consider different available lines of justification of the principle. I defend the view that there is enough in those arguments to at least conclude in favor of a certain

conditional form of validity for the principle: the principle is valid in all those situations where communication and genuine interpretation are actually achieved. This conditional validity is all I need, and could hope to secure from this discussion, to support my own thesis.

The general and foundational account of meaning in the background is one that takes as its cornerstone the publicness of meanings. If meanings are to be public and sharable, they cannot be beyond the reach of an ordinary speaker's ability to discover them, in other words, meanings must be determined by interpretability. This position, however, is not without its challenges. Despite the strong intuitive appeal of the claim that linguistic meanings ought to be public and sharable, the idea of subjecting the reality of meanings to the interpreter's ability to discover them seems bound to clash with other equally strong intuitive convictions. I will address two instances of this sort.

First, I accept the point that endorsing the view I am defending should not demand that we give up on the idea that speakers have direct - non-inferential, i.e. non-interpretational - and authoritative knowledge of their own meanings and thoughts. In the first part of Chapter 3, I argue that there is a Davidsonian middle path that privileges neither the speaker's (first-person) nor the interpreter's (third-person) stance but, instead, advocates the need for an agreement and interdependence between perspectives in the determination of the intentional and semantic facts.

Second, and obviously related, I resist the objection that the interpreter's stance is too dilapidated to be able to concur with the speaker's perspective over her meanings. In the second part of Chapter 3, against the idea of abundant, and easy to prove, *underdetermination* - and, perhaps, *indeterminacy* - of interpretation, I stress the often underestimated power and scope of the Principle of Charity.

Finally, in Chapter 4, I turn to the empirical study of language acquisition. I return to the point where I left the discussion at the end of Chapter 2, having concluded in favor of the conditional validity of the Principle of Charity, and attempt a scientifically informed follow-up. I endeavor to show - in some detail and with reference to a few experimental studies - that assumptions of rationality, coherence and correspondence are not only licit as strategies or steps in interpretation, but that they are actually adopted and explored by children in the course of language acquisition and, more specifically, word learning. I report that this state of affairs is generally acknowledged in the field, i.e. that there is strong agreement around the idea that mindreading and rationalization is a crucial element in the process of word learning.

1. Davidson's Conception of Meaning and Interpretation

After a brief introduction to Davidson's program of truth-theoretic semantics, we will begin the exploration of Davidson's conceptual experience of Radical Interpretation. I will present the method through which an interpreter, according to Davidson, should be able to acquire knowledge of the speaker's meanings and, simultaneously, of her propositional mental states. The question of linguistic interpretation, as we will see, grows naturally into the question of general interpretation.

The Principle of Charity, a key piece in this entire dissertation as well as in Davidson's work, will make its first appearance. I detail its role in the inquiry developed by the interpreter.

Davidson's views on Radical Interpretation both imply and follow from a general conception of meaning, communication and language, a conception that - in its main lines - I will be endorsing and defending throughout this dissertation. In the last section, I develop and clarify the main features of this view.

1.1. TRUTH-THEORETIC SEMANTICS

There is a distinction that is frequently invoked in the philosophical study of meaning. Stalnaker (1997, 535) coined the labels '*descriptive semantics'* and '*foundational semantics'*, but a clear idea of this contrast had been around at least since David Lewis' *General Semantics* (1970). A descriptive theory of meaning is one that specifies the meanings, or semantic contents, of all expressions of a certain language. A foundational theory is one trying to reveal the underlying facts - sociological, psychological, neurological, biological, or whatever - making it the case that a certain community or person is using a certain language

or idiolect[3].

Radical interpretation is the conceptual experience Davidson uses to explore the foundational line. My main interest lies there and it will occupy me from the next section on.

Davidson (1967) proposes a *truth-theoretic semantics* for the descriptive role. That was his earliest and, arguably, most influential proposal in the study of language and meaning. For the purpose of this chapter, and of this dissertation, I need not go into its many details, theoretical or formal, let alone evaluate its adequacy or merits. Nonetheless, a short introduction is in order and this is what I will dedicate this first section to.

Davidson conceives an adequate descriptive semantics for a language *L* to be one specifying something knowledge of which would be enough for an interpreter to be able to understand the utterance of any sentence of *L*. What kind of statement could count as providing enough for the understanding of a sentence *s* of *L?* According to Davidson, a possible answer would be: a specification of the truth-conditions of *s, stated as a T-sentence*, belonging to a Tarskian truth theory for *L* satisfying a number of conditions.

Davidson's central insight was that the purpose of a descriptive meaning theory would be best served if one found a way of associating the sentences of the language to their meaning specifications, not by describing or referring to propositions or meanings, but simply by presenting those meanings *in use*.

All that is needed then is to find some way of associating each and all the sentences of the language with their respective meanings. The theory could not simply consist of a list of each particular association. Such a list would be infinite given the infinite number of sentences belonging to each natural language. Besides unpractical, a listiform theory would be unsuited as a representation of a speaker's competence. Speakers are finite beings and, accordingly, Davidson intends to capture their competence in a finite form (Davidson, 1965).

Two properties are evident in language and linguistic competence - *creativity* and *systematicity*. Creativity manifests itself in the fact that every natural language is such that an infinite number of new sentences, sentences never before formulated, belong to it. Any speaker of the language will understand those new sentences on the basis of the same competence that allowed her to understand the previous sentences she had used before.

[3] Recently, "*Metasemantics"* seems to have taken over as the new prevailing label for this type of investigation. See Burgess and Sherman, 2014, Introduction.

Systematicity has to do with the recurrence of identical parts in different expressions, always carrying the same semantic value. Accordingly, speakers will make use of the same *piece of competence* in dealing with them.

A third property, *compositionality*, explains these first two: it is because the meanings of complex expressions - namely sentences - are determined by the meanings of their component parts and the way in which those parts are combined that natural languages exhibit both creativity and systematicity. Davidson sought a descriptive semantics that captured the compositionality of language.

He settled on a formal theory - a truth theory or T-theory - that, starting from a finite number of axioms, was capable through a method of proof of generating an infinite number of meaning-specifying theorems. For this purpose, Davidson explored Tarski's work (1933, 1944) on the definition of the truth predicate for formal languages. Tarski had developed a method for determining the extension of the predicate '*true-in-L'* as applicable to the sentences of any formal language that satisfied a certain number of requirements.

He had shown how to arrive at a T-sentence for every sentence of L, starting only from a finite number of axioms specifying the semantic value of each primitive expression of L and the values of complex expressions of L calculated on the basis of its simpler ingredient expressions.

Tarskian T-sentences, Davidson noticed, meet the condition set out above of linking sentences of L with a specification of their meanings in use - they specify the meaning in the language in which the theory is stated, the metalanguage. Take a sentence *s* of *L*, referred to by some singular term, preferably some structural description, and a specification *p* of its meaning and truth-conditions, and combine the two in the same sentence according to the following scheme: *s* is true in *L* if and only if *p.* The result will be what Davidson calls a T-sentence for *s.* A simple example would be: 'The cat is on the mat' is true in English if and only if the cat is on the mat.

In this way, a Tarskian truth theory for a language L seems capable of doing double duty as also a descriptive semantics for that language. Starting from finite resources, the theory is able to produce a theorem relating each sentence of the language with its meaning.

Even in a brief and simplifying summary such as this, a few words about the difficulties facing this project must be put forward. First, Tarski had not been interested in

natural languages. He had shown how to define a truth predicate as it applied only to formal, well-groomed and well-behaved languages. A lot of work and ingenuity has been spent in the last few decades trying to adapt this approach to natural languages - Lepore and Ludwig distill some of the best of it in (Lepore & Ludwig, 2007). Opinions vary regarding the successes of this program.

The second difficulty has to do with Davidson's reluctance to uncritically accept an ordinary notion of meaning and the commonsense view of synonymy and translation. Tarski had no question with meaning. He was only after the notion of truth, and he felt free to assume an unproblematic specification/translation of the primitive terms of the language into the metalanguage. Davidson did not take the same option.

Assuming interpretive axioms would be the easiest way to ensure the ensuing theorems would also be interpretative (Lepore & Ludwig, 2005, 72ff). That is, assuming interpretive axioms would be the easiest way to ensure that the sentence on the righthand side of the biconditional in every T-sentence actually specified the meaning of the sentence referred to on the left. In refusing this option, Davidson must put something else in its place. The form of the theorem itself only requires co-variance of truth-values between s and p; it does not ensure that the second interprets the first.

> There is no difficulty in rephrasing Convention T without appeal to the concept of translation: an acceptable theory of truth must entail, for every sentence *s* of the object language, a sentence of the form: *s* is true if and only if p. where 'p' is replaced by any sentence that is true if and only if *s* is. Given this formulation, the theory is tested by evidence that T-sentences are simply true; we have given up the idea that it must also tell whether what replaces 'p' translates *s*. It might seem that there is no chance that if we demand so little of T-sentences, a theory of interpretation will emerge. And of course this would be so if we took the T-sentences in isolation. But the hope is that by putting appropriate formal and empirical restrictions on the theory as a whole, individual T-sentences will in fact serve to yield interpretations. (Davidson, 1973, 135)

Instead of assuming interpretive axioms Davidson tries to make do with alternative conditions and empirical tests not involving an appeal to semantic notions more problematic than truth. He considers various constraints that, especially when working together, seem indeed to add some plausibility to the interpretive character of the qualified T-theory.

First, there is Davidson's emphasis on the holistic constraint set on the theory (Davidson, 1967, 26). It would be easier for a true T-sentence to fail to be interpretive when

considered in isolation. The fact that the theorems must be the product of a theory generating true T-sentences for each sentence of L limits very significantly the range of options for axioms and theorems. Then there is also the added empirical traction resulting from T-sentences dealing with indexical and demonstrative sentences (Davidson, 1967, 35). Natural languages have richer links to the world than the formal languages considered by Tarski, richer links that the T-theory is expected to take into account. A third constraint added later to the others, albeit a non-extensional one, is that the theorems should have the force of laws, that is, should not only be true but nomologically so (Davidson, 1984d, xiv).

With these innovations and the emphasis on the systematic interrelations between theorems, Davidson did manage to show how an ingenuous exploration of the concept of truth can bring us closer to meaning and interpretation than we might have initially believed by attending merely to the form of the theorems. Yet it is still far from certain that the theory has all that it takes to count as a genuine meaning theory. The last move to consider is radical interpretation. Davidson's conceptual experience - a topic of investigation in itself, responding to its own specific philosophical drives and concerns - contributes also, and crucially, to the completion of Davidson's proposal on descriptive semantics.

As Lepore and Ludwig put it, after having first attempted to defend that "a merely extensionally adequate truth theory for a natural language (i.e. one that is simply true) would thereby meet Tarski's Convention T or an analog for natural language" (Lepore & Ludwig, 2005, 75) - that is, would generate interpretative theorems - in a "second stage, when it became apparent that this condition was too weak, he appeals to confirmation by the procedures of a radical interpreter as an additional constraint" (ibid.). The method and spirit presiding over radical interpretation - to which, shortly, I will turn my attention - can be expected to add the fine-tuning that is eventually still missing. This is also Kathrin Glüer's understanding of the case (Glüer, 2011, 65). Davidson himself does not declare *the move* in so many words, but the solidarity between his proposals on truth-theoretic semantics and radical interpretation is beyond doubt.

We now have a clearer idea of the knowledge Davidson expects his interpreter to acquire about the speaker of a language unknown to him, and we have learned about the form of the theory in which that knowledge is supposed to be captured. I am afraid this improvement is only tenuously relevant for my main interests and purposes. However, given the occasional need to refer to theories, theorems, proofs, T-sentences and the like, I decided

to include this minimal background account.

1.2. RADICAL INTERPRETATION

Let us move on to the second topic of reflection - the foundational issues. What facts underlie the semantic values of sentences and utterances? How could a radical interpreter explore them so as to arrive at adequate ascriptions of meanings and thoughts?

1.2.1. THE LINKS BETWEEN LANGUAGE AND THE WORLD

The key for any interpretation project must be found in the correlations between the uses of linguistic expressions and the circumstances accompanying those uses. There are different types of knowledge one can have of, and get from, these correlations. In order to get to these differences, let me introduce a quasi-technical notion that I will be using a great deal - *hold true. Hold true* is the attitude of taking some declarative sentence to be true, in whatever form or circumstance. In fact, it is an ordinary and natural notion, the only quasi-technical thing about it being that it will be used in a markedly general and intensive fashion to encompass all other more specific attitudes of acceptance relating individuals to sentences. There are many ways of holding a sentence true. One can simply affirm it sincerely, or assent to someone else's utterance of it, but one can also do so by giving the right sort of answer to a certain question, by keeping quiet, by sighing; in sum, by all sorts of signals and actions.

The first moment of radical interpretation is all about establishing the right correlations between held true sentences and the circumstances in which they are held true. The idea behind it is that, in the context of radical interpretation, the recognition that a certain sentence is held true is easier than and prior to the actual understanding of the sentence, and hence that the interpreter must explore this more accessible knowledge, searching for how it can be related to what else is going on on his path to full interpretation.

In a sense, Davidson simply grants the interpreter the ability to recognize situations where the foreign speaker holds true some particular sentence. This seems plausible enough. Even without understanding her language or possessing detailed knowledge about her mental states, the interpreter can be expected to recognize - certainly not at first sight, but with little enough effort - when the speaker is asserting something, or when she is responding positively

or negatively to a certain query.

However, being able to find out that a sentence is held true on certain occasions is only of any use to the interpreter if, subsequently, there is a way to get from there to the learning of the sentence's actual meaning. In a first moment, the interpreter should be focusing only on a special brand of sentence, those sentences to which assent is made fully pertinent by what is patent in the circumstances prompting that assent. Let us call them *perception sentences*[4].

Consider three types or levels of correlation between sentences and the occasions on which they are held true. First, we could be talking about a mere *statistical correlation* between the holding true of the sentence and the surrounding circumstances, independently of any further relation that might hold between them.

Second, we could have a little more than mere concomitance. It could happen that some particular set of circumstances regularly accompanying the holding true of a sentence were also *causing* it in some way.

Third, we could be talking about a still more intimate relation between sentence and occasion. In this case, the circumstances accompanying and causing the holding true of the sentence also *give it its content*. This is the kind of correlation that matters to us, the kind that marks the distinctive kind of sentence we called perception sentences. Any project of interpretation from scratch must start with these sentences whose content is, in some sense, *readable* from the occasions on which they are held true.

1.2.2. THE ACCOMPANYING CIRCUMSTANCES

In 'radical interpretation', the qualification 'radical' refers to what is meant to be the fundamental character of the experience. The project is that of describing a possible process of interpretation of a speaker when starting from an initial situation of complete ignorance of her language. Another way to put it is to say that the project is that of looking for a

[4] The expression is taken from Davidson (2001e, 138). I pick it because I am not satisfied with an undifferentiated appeal to occasion sentences for this job as Davidson sometimes opts for. There are differences between occasion sentences that are worth marking. Even if no sharp border can be found, there are occasion sentences whose content is much more *observable* or *perceptual* than others - those are the 'perception sentences'.

justification of an interpreter's understanding of some speaker's words - i.e. a justification for his beliefs about the meaning of such words - without, naturally, already presupposing that he possesses such knowledge.

Fitting in with what was said in the previous section, Davidson sustains that, in radical interpretation, "the evidence available is just that speakers of the language to be interpreted hold various sentences to be true at certain times and under specified circumstances." (Davidson, 1973, 135) This, of course, invites a very important question: How are we to describe the circumstances accompanying the holding true of a certain sentence? What kind of recognition and description of the circumstances is the interpreter allowed?

The first, obvious, condition is that these descriptions must not involve the actual knowledge we are trying to justify. But what exactly is the knowledge that he is trying to justify? Davidson is not very clear about what, or how much, is being justified and what, or how much, is allowed to be taken as granted.

For instance, it would sabotage the whole project to try to have started it with observations such as: *"The speaker holds the sentence 'Le chat est sur le tapis' true whenever he intends to mean that the cat is on the mat"*, or *"The speaker uses affirmatively the sentence 'le chat est sur le tapis' when he believes that the cat is on the mat."* First of all, and most importantly, such facts are simply not openly there for an interpreter still ignorant of the language to observe. Second, one would not be able to tell apart observations so detailed and informed as these from the actual pieces of knowledge that the radical interpreter is expected to discover and confirm on its basis. It would amount to no learning, discovery or justification beyond mere observation. How could one radically interpret? Can one justify one's knowledge of other speakers' meanings? '*Just observe them!*' would be a disappointing answer.

However, the danger of failing to meet that first condition does not come from cases such as these where the nullity of the explanation and justification results are so perfectly flagrant. The real concerns should arise instead with respect to descriptions where the presupposition of the very knowledge being confirmed is much more subtly installed. In particular, we should pay attention to the danger of taking for granted too much about the speakers' psychology, about their language, about the way they live, experience their environment and relate to each other. Some assumptions will indeed be needed, even if tested

afterwards and not simply taken for granted. The question is where to draw the line. How can we ensure, simultaneously, that the interpreter will have enough information at his disposal to render interpretative success plausible, and that his results will not be insufficient or null - that is, that some real progress in understanding is actually achieved instead of just smuggled in with the observations? What type of descriptions can do the trick? We will be returning to these problems and questions, and elaborating on their answers, time and again throughout this dissertation. Answers are complex and will be fully disclosed only gradually.

Here is one more crucial element, one further determination, this time a positive one: the description of the circumstances accompanying the holding true of a perception sentence must be made in terms of the salient distal features of the scene shared by speaker and interpreter.

> The distal theory, on the other hand, depends primarily on shared causes which are salient for speaker and interpreter, learner and teacher. Meanings are shared when identical events, objects or situations cause or would cause assent and dissent. As a radical interpreter I correlate verbal responses of a speaker with changes in the environment. Inferring a causal relation, I then translate those verbal responses with a sentence of my own that the same changes in the environment cause me to accept or reject. This is the distal theory at its simplest, subject to various fairly obvious caveats. (Davidson, 1990b, 54-55)

Davidson picks *distal stimulus - events, objects or situations* - as the causes of the holding true or false of perception sentences. The notion of a distal stimulus is to be contrasted with that of a proximal stimulus. As we will see, Quine, in his approach to the same task of radically interpreting (in his case, translating) the words of another, favors proximal stimuli. This means that he will be taking the relevant causes of linguistic behavior to be the stimulations of sensory receptors. Sensory stimulations are called *proximal* because they are located as close as possible to the speaker, on the surface of his own body, in his nerve endings. The distal cause is separated from the speaker’s body and belongs to the public and objective environment that speaker and interpreter actually share.

What qualifies as a distal stimulus? First and foremost, all the common objects and events populating our natural worldview, all those things we spontaneously notice and talk about, seem to qualify. But that does not close it. If I have got it right, everything that can work as an objective and shared (or shareable) cause of perceptual agreement between speakers can work as a distal stimulus.

An option for distal stimulus implies the need for some type and degree of similarity between speaker and interpreter in the way they think and experience the world and their shared environment. What we have here is exactly the type of tension anticipated above, on the condition that the description of the circumstances should not presuppose too much about the speakers. However, it will become clear that starting at a distal level does not compromise the account's worth. There is still much to be learned even after assuming some general common ground, and hence still much learning to be depicted and explained.

1.2.3. FROM OBSERVATIONS TO A MEANING THEORY

What is the interpreter to do with the observations he manages to gather? How is he to continue the interpretation process? According to Davidson, first, he must inductively generalize them; then, he will transform those generalizations into truth conditions and meaning ascriptions; lastly, he must analyze the base that is made up of these first perception sentences to work out a full meaning theory. Let us follow this in more detail.

He starts with observations such as:

(E) Kurt belongs to the German speech community and Kurt holds true 'Es regnet' on Saturday at noon and it is raining near Kurt on Saturday at noon. (Davidson, 1973, 135)

This sort of observation must be accumulated. In this process it is natural for the interpreter to gradually arrive at sharper and finer-grained descriptions of the relevant features of the circumstances of utterance, confirming the constant presence of some features, and infirming that of others. Induction will allow inferences of statements such as G.

(G) (x)(t) (if x belongs to the German speech community then (x holds true 'Es regnet' at t if and only if it is raining near x at *t*)) (ibid.)

Then follows an important leap - the passage from observable correlations between uninterpreted linguistic behavior and the circumstances in which they take place to the ascription of content to the speaker’s utterances and thoughts. In other words, the passage from the statement of the *conditions in which a certain sentence is held true* to the statement of *its true conditions*.

(T) 'Es regnet' is true-in-German when spoken by *x* at time *t* if and only if it is raining

near *x* at *t.* (ibid)

At this point, T is but a *proto-theorem* - to give it a name - that is, a statement that already possesses the form the theorems of the theory should have and indicates the truth-conditions of the sentence in question. It differs from a real theorem only in not having been deducted from the theory's axioms by a proof method. That is, of course, due to the fact that the interpreter still does not have a theory.

The interpreter must then go on trying to gather as many proto-theorems as possible. He will use them as targets for the theory to be constructed.

The next step is to try to *read* a logical structure in the foreign sentences - a logical structure adequate to a theory that is to be built in accordance with Tarski's model. The interpreter must go through a process that is, in a sense, the reverse of the proving of a theorem. Instead of starting with the axioms and progressing deductively towards theorems, he must start with the proto-theorems and seek a set of axioms capable of generating them. Once he finds such a set, he will not only be capable of generating the actual theorems corresponding to the proto-theorems he started with, but also, of course, a potentially infinite number of other theorems, about other sentences, built out of the same lexicon.

The new theorems will provide new tests for the theory. The interpreter must return to the observation stage to confirm whether or not the theory's previsions are satisfied empirically. If not, modifications and readjustments are called for until the interpreter can be satisfied with a sufficient degree of agreement between observations and previsions.

In all, we can distinguish five main stages in the process:

1. observation/test
2. generalization
3. proto-theorems
4. logical analysis and theory building
5. new theorems and predictions (after 5, the interpreter must return to 1.)

These are the broad outlines. Now we must pay more careful attention to what is involved in some of these stages. Stage 3 is crucial, as it consummates the passage from what is open for the interpreter to see to the first content ascriptions - retrospectively, however, the

previous steps are not at all innocent of anticipated understanding and content shaping, as will be emphasized later. How is this crucial move from observational reports to ascriptions of truth-conditions achieved? What is sanctioning it?

1.2.4. THE PRINCIPLE OF CHARITY

The holding true of a sentence by a speaker is the product of two factors: the meaning she ascribes to the sentence, and a certain belief of hers. The speaker holds *s* true because she believes that *p* and takes *s* to mean that *p*. If the interpreter knew the meaning of her words, her utterances would reveal her beliefs to him. Inversely, if he knew her beliefs (and further mental states) sufficiently well, he could infer the meaning of her words. The thing is that he ignores both. What he needs is a way of breaking this circle, of unveiling both her meanings and her thoughts.

> Since we cannot hope to interpret linguistic activity without knowing what a speaker believes, and cannot found a theory of what he means on a priori discovery of his beliefs and intentions, I conclude that in interpreting utterances from scratch—in radical interpretation—we must somehow deliver simultaneously a theory of belief and a theory of meaning. (Davidson, 1974a, 144)

Davidson's proposal is "to solve the problem of the interdependence of belief and meaning by holding belief constant as far as possible while solving for meaning" (Davidson, 1973, 137). He explains that this is to be "accomplished by assigning truth conditions to alien sentences that make native speakers right when plausibly possible, according, of course, to our own view of what is right." (ibid.) To conform to such a plan is to adopt the *Principle of Charity*.

The plan is supposed to get the interpreter out of the impasse just described. First, the interpreter is now meant to "take the fact that speakers of a language hold a sentence to be true (under observed circumstances) as prima facie evidence that the sentence is true under those circumstances" (Davidson, 1974a, 152). However, being allowed to presume the sentence uttered, and the correspondent belief, to be true is hardly enough for interpretation. Naturally, the plan calls for more. Davidson expects the interpreter not only to take the native sentences as true, but he also finds him in a position to assign those sentences, and the correspondent beliefs, a proper content - that is, true conditions.

Exactly how? It is still licit to ask. The idea, only implicit here - although fairly

obvious, and to be clearly formulated in later texts - is that the interpreter must *read* those truth conditions from what is going on at the time and place in which the sentence is held true. Such reading will necessarily involve either, or both, some form of picturing oneself in the native's shoes or the possessing of good enough knowledge about what one is likely to think and say when one is wearing them. Assuming the sentence in question is a perception sentence – and so the expression of a belief elicited by the occasion on which it is held true, and grounded on what is patent there – besides presupposing the speaker to be correct, the interpreter will further need either to be able to project himself into the speaker's perspective or to know enough about his general psychology and worldview. Only one of those, or a mixture of both, would get him from truth to actual content.

The presupposition of truth provides an occasion propitious to the discovery of the right link between utterance and circumstances of utterance. By *right link* I mean, of course, the one determinant of the sentence's meaning. To complete the deed, the interpreter still has to be able to recognize the link in question. After assuming that the speaker is right in the adequacy she – indirectly, through her assertive behavior – claims to exist between utterance and occasion of utterance, he must still assume some sort of agreement between him and the speaker, in virtue of which he is apt to grasp precisely that same link between utterance and occasion.

What is at stake here is Davidson's Principle of Charity, in both its main dimensions distinguished by Davidson in a late text (Davidson, 1991, 211) - *the Principle of Coherence* and *the Principle of Correspondence*[5]. The Principle of Coherence tells the interpreter to assume that an agent's thoughts, meanings and actions cohere - logically, conceptually, inferentially - in a number of interesting and complicated ways. This component is involved in the interpreter's employment of the equation presented above relating meaning, belief and a sentence held true: the interpreter must detect a sentence held true, figure out the relevant belief, and thus infer the meaning of the sentence. In figuring out the relevant belief, it is the

[5] It is only rather late in his work - as far as I can tell, only in 1991, in "Three Varieties of Knowledge" - that Davidson clearly distinguishes these two principles and gives each a name of its own. However, as early as 1967, in "Truth and Meaning", we can already perceive those two strands running their separate yet intimately related ways: "The linguist...will attempt to construct a characterization of truth-for-the-alien which yields, so far as possible, a mapping of sentences held true (or false) by the alien on to sentence held true (or false) by the linguist. . . . Charity in interpreting the words and thoughts of others is unavoidable in another direction as well: just as we must maximize agreement, or risk not making sense of what the alien is talking about, so we must maximize the self-consistency we attribute to him, on pain of not understanding him" (Davidson, 1967, 27).

second component, the Principle of Correspondence, that is crucial.

1.2.4.1. THE PRINCIPLE OF CORRESPONDENCE

Davidson explains that the Principle of Correspondence “prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances” (ibid.), and so endows the speaker “with a degree of what the interpreter takes to be true belief about the world” (ibid.). This is what is supposed to allow the interpreter to read the relevant beliefs from the speaker’s exposition to their shared environment.

In their 2005 book, Lepore and Ludwig (L&L) assess “three interpretations of the principle of charity intimated by various passages in Davidson’s work” (Lepore & Ludwig, 2005, 175ff.). They use the generic label “the principle of charity” but then they explain that “Davidson usually treats under the heading of ‘the principle of charity’ a family of principles, not all of which are directly relevant to providing a warrant from inferring TF-sentences from L-sentences” (ibid.185)[6], making it clear that they are only dealing here with the *correspondence* component of the principle.

L&L find all three conceptions insufficient, and consequently they advance a fourth proposal that they claim to be apt to play the role the principle was introduced to play - that of securing the passage from observations and generalizations to actual ascriptions of interpretative truth conditions to the foreign sentences.

The first one they call *Veracity*.

(Veracity) For all speakers *S*, times *t*, sentences *s*, *ceteris paribus*: *S* holds true *s* at *t* iff *s* is true(*S*, *t*). (ibid., 186)

They show that Veracity does not provide the kind of support the interpreter needs at this point. This principle goes no further than constraining the interpreter to interpret the speaker’s sentences so that they turn out true (or false) when the speaker holds them true (or false). This, as I explained above, is not enough to warrant interpretation. We need something more to get from truth to the actual content or truth conditions of those sentences. With

[6] By *‘TF-sentences’* they mean the type of proto-theorems I exemplified above with T. By *‘L-sentences’* they mean the type of generalizations I exemplified above with G.

veracity in place we would still be allowed to build all sorts of extensionally correct truth theories delivering all sorts of true but non-interpretative theorems.

They try a second interpretation they call *Charity.*

(Charity) For any speaker *S*, time *t*, belief *b*, *ceteris paribus*: *b* is a belief of *S*'s at *t* about and prompted by *S*'s environment iff *b* is true. (ibid., 189)

While the first formulation, Veracity, concerns beliefs towards sentences - stating that the speaker must be right in holding true sentences as observed by the interpreter - the second, Charity, concerns *simple* beliefs - stating that the speaker must be right in believing what he believes on such occasions - those beliefs presumably determining the holding true of the sentences. The difference between them is not very relevant to my concerns. They are in no way incompatible, and it seems clear to me that Davidson means the principle he advocates to comprehend both claims and more.

Being so closely related, this second attempt fares no better than the first and it promotes no advance towards interpretation. The interpreter is still stopped at truth, with no clue about how to get to content.

L&L try a third principle they call *Agreement.*

(Agreement) For any speaker *S*, time *t*, belief *b*, *ceteris paribus*: *b* is a belief of *S*'s at *t* about and prompted by *S*'s environment iff *b* is true. (ibid., 190)

L&L affirm that "Charity requires Agreement" (ibid., 191) but that Agreement does not require Charity. I concur on the second but differ on the first claim. Indeed, they are right as Agreement does not imply Charity because Agreement allows the speaker to be wrong in his beliefs, as long as the interpreter is agreeingly wrong, whereas Charity necessitates the speaker to be right. As L&L put it, "we can squeeze Charity out of Agreement only by assuming the interpreter's beliefs are mostly true" (ibid., 191).

However, somewhat surprisingly, the exact reverse is also the case: *we can squeeze Agreement out of Charity only by assuming the interpreter's beliefs are mostly true*. Charity says nothing about the interpreter being right. We would get no agreement if the speaker were right, as Charity prescribes, and the interpreter were wrong, as Charity allows.

From the point of view of the interpreter applying the principle, again both interpretations amount to the same. If the interpreter is following Charity he will be looking

for true beliefs according to his "own view of what is right" (Davidson, 1973, 137), but if he is following Agreement, again he will be looking for beliefs that he takes to be true - otherwise he would not believe them.

This new interpretation is still not what we are looking for - L&L defend this and I agree. Agreement is insufficient for interpretation for the same reasons that the previous two interpretations did not work: it is only about sameness of truth-value and neglects sameness of meaning or truth-conditions.

As I have said above, L&L propose a fourth interpretation, *Grace*. I find their complex exposition mildly puzzling so I quote abundantly before I risk my reading of what is going on. Below is Grace.

> (Grace) *Ceteris paribus*, when we replace '*p*' in (S)
>
> > (S) *S* believes at *t* that *p*
> >
> > with a sentence that expresses the content of an environmentally prompted belief of *S*'s, the sentence expresses also a condition in *S*'s environment that prompts that belief. (Lepore and Ludwig, 2005, 194)

In other words, as I understand it, Grace asserts that for any belief belonging to the particular class of beliefs[7] that radical interpretation must start with, the same sentence that expresses its content also expresses the condition in the believer's environment that prompts the belief.

L&L claim that this new (interpretation of the) principle finally suffices, together with *two other assumptions*, to get the interpreter from observations to interpretative truth conditions. Below are the assumptions:

> (1) *Ceteris paribus*, the *L*-sentences that the interpreter confirms identify conditions under which a speaker holds true sentences and these conditions are also the prompting conditions, if any, which those beliefs, that are the basis on which the speaker holds those sentences true, are about.
>
> (...)
>
> (2) *Ceteris paribus*, a speaker *S* holds true *s* at time *t* because, and only because, he knows that *s*

[7] What we now - in an allusion to their relation to the sentences I have called perception sentences - may call 'perception beliefs' - beliefs not only *environmentally prompted* but also *environmentally determined,* or *specified*.

> means(*S*, *t*) that *p* and believes that *p*, and knows that if *s* means(*S*, *t*) that *p*, then *s* is true(*S*, *t*) iff *p*. (ibid.194-5)

To show that they are enough for interpretation, L&L start with a schematic generalization,

> *S* holds true *s* at *t* iff *p* (ibid., 195)[8]

and proceed by explaining their reasoning. First, since some belief always has to contribute for the holding true of any sentence,

i. 'we know that some belief is the basis of S's holding true s at t.' (ibid.)[9]

Then Grace is called to do its work: it allows the interpreter to take *s* to be about what in the surrounding environment caused the speaker to hold it true,

ii. 'Grace tells us the belief on the basis of which *S* holds true *s* is about a prompting condition;' (ibid.)

However, that is not enough. It could very well happen that the content of *s* was determined by the prompting condition and yet the interpreter was unable to grasp the right condition in his attempted generalizations. That is why we need assumption (1): we must assume that the interpreter manages to get to the right generalizations - i.e. that he manages to recognize, as I formulated above, the right link between utterance and circumstances of utterance, the one determinant of the sentence's meaning.

iii. '(1) tells us that '*p*' expresses the right condition.' (ibid.)

Now it is time for the second assumption. Assumption (2) is quite different in nature from assumption (1). On the one hand, (2) is a much more general claim, in that it has applications beyond the filing of this particular gap between observation and interpretation. For instance, every speaker has to know it, in some sense, in order to be able to rationalize her own holding true of some sentence. On the other hand, (2) is a much more trivial claim. It presents a piece of common wisdom about language and language use that no speaker fails to grasp, even if only implicitly in most cases.

[8] For simplicity's sake, I have removed the asterisks and the '0's that figure in the original passage.
[9] I added the listing - i., ii., ….

Assumption (1) is a very different matter. It seems much more limited in its application - it concerns only the ability to radically interpret others - and it is an essential ingredient to such an ability. In this sense, it should probably be taken as actually belonging to the principle of correspondence itself, not as a foreign assumption. Also, contrasting with the trivial truth of assumption (2), the validity of (1) seems perfectly challengeable. Its eventual truth requires argument and justification – we will see if some can be provided in the sections ahead about the justification of the principle of charity.

Returning to L&L's argument, assumption (2) tells us that the sentence the speaker holds true expresses the content of his belief,

iv. (2) tells us that *s* expresses the content of that belief relative to *S* at *t*. (ibid.)

Now, since *p* specifies the condition that the belief is about (according to ii and iii), s expresses that belief (according to iv), and a speaker will hold a sentence true if and only if that sentence's meaning is identical with the content of one of his beliefs (according to assumption (2)),

v. we can infer, *ceteris paribus*, *S* holds true *s* at *t* iff *s* means(*S*, *t*) that *p*. (ibid.)

From v. we can finally infer the truth of an interpretative ascription like (3),

vi. But then, *ceteris paribus*, (3) is also true, where (and because) the sentence on the right interprets the one on the left.

(3) *s* is true(*S*, *t*) iff *p*. (ibid.)

Here is another simpler but reasonably close version of this argument. The idea is to start with the statement of some generalization, a, and to end with the statement of its correspondent interpretative truth-conditions.

a. *S* holds true *s* at *t* iff *p*

Then, Grace together with assumption (1) tell us that the belief in virtue of which S holds s true is about the condition identified by the interpreter as prompting that holding true. So, we can write b,

b. *S* believes that *p*

Now, assumption (2) tells us that the meaning of what S holds true is identical to the content of the belief responsible for his holding it true. This gives us c,

c. *s* means that *p*.

Given the relationship between the meaning of a declarative sentence and its truth-conditions, from c we can deduce d,

d. s is true(*S*, *t*) iff p

Compared with the previous three interpretations of the principle of charity, there are two crucial differences:

A. Now the interpreter is supported in taking the speaker's beliefs and words to be about those same conditions that prompt those beliefs and words - thanks to Grace.

B. Now the interpreter can assume that he is apt to identify what those conditions are - thanks to assumption (1).

L&L seem (at least) reluctant to ascribe to Davidson the acceptance that Grace, or something tantamount to it, is necessary to bridge the gap between observation and generalization to content ascription. For instance, they say that "Grace is a substitute for the principle Davidson actually invokes" (ibid., 195). They concede that "in work after *Inquiries*, Davidson has argued it is the upshot of reflection on the radical interpreter's procedures that a speaker's beliefs are about conditions in his environment they are nomically correlated with" (ibid.), but again they hesitate and declare that "it is not clear, however, whether Davidson realizes Grace does not follow from Charity, and is essentially stronger" (ibid).

I do not share this reluctance. I find it fairly evident, at least in his later texts, that Davidson acknowledges the necessity of such a principle, and includes something akin to Grace in his articulation of the principle of charity. See, for instance, the following passage, part of which was already quoted above,

> The Principle of Coherence prompts the interpreter to discover a degree of logical consistency in the thought of the speaker; the Principle of Correspondence prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances. Both principles can be (and have been) called principles of charity: one

> principle endows the speaker with a modicum of logic, the other endows him with a degree of what the interpreter takes to be true belief about the world. (Davidson, 1991, 211)[10]

Davidson is explicitly talking about the principle of charity. To say that the principle "prompts the interpreter to take the speaker to be responding to the same features of the world that the interpreter would be responding to" is just another way - admittedly synthetical - to state A and B above, that is, to admit Grace and assumption 1.

In other places, Davidson develops what he means by *responding to features of the world* - for instance, in "A Coherence Theory of Truth and Knowledge" Davidson explains that

> … we can't in general first identify beliefs and meanings and then ask what caused them. The causality plays an indispensable role in determining the content of what we say and believe. (...)
>
> … we must, in the plainest and methodologically most basic cases, take the objects of a belief to be the causes of that belief. And what we, as interpreters, must take them to be is what they in fact are. Communication begins where causes converge: your utterance means what mine does if belief in its truth is systematically caused by the same events and objects. (Davidson, 1983, 151)

In the end it seems safe to affirm that Davidson - even if after a hesitant first start - fully realizes the role that the principle of correspondence must play in the methodology he designs for his radical interpreter, and that he puts it forward adequately in some of the formulations he provides.

Henceforth, whenever I refer to the *Principle of Correspondence*, I refer to a claim that includes Lepore and Ludwig's *Grace* and *Assumption 1*. It asserts that speaker and interpreter share a sufficiently similar and veridical recognition of their surrounding environment and circumstances in such a way that the interpreter is frequently able to detect when the speaker's utterance is caused and is about some commonly perceived feature of their shared situation.

Eventually, I will explicitly withdraw from *Correspondence* into mere *Agreement*[11]. Affirming Agreement is the same as affirming Correspondence except for the emphasis on the veracity of the shared recognition of what is going on. This will happen, in particular, in

---

[10] See also, for instance, Davidson, 1984a, 35-6.

[11] Which must absolutely not be confused with Lepore and Ludwig's version of the principle, quoted above, that went by the same name. I must risk some confusion here because there is no equally good label for the weaker form of Correspondence that I am trying to grasp here.

the context of the justification of the principle. There I retreat not because I find it indifferent whether or not speaker and interpreter are both right or wrong in their beliefs, nor because I have any reason to doubt that they will actually be right most of the time. It is just because I decline to engage with the skeptical challenges that the emphasis on truth appears to commit one to.

1.2.4.2. THE PRINCIPLE OF COHERENCE

As I have already explained above, whereas the first component of the principle of charity, correspondence, concerns *external* relations - between thought and the world, and between different thinkers - the second component, coherence, concerns *internal* relations between thoughts - relations of logical (in a broad sense) consistency between thoughts and their contents in the mental life of each individual thinker. Some of them feel so natural and evident that they tend to run very silently. That, of course, makes them no less essential. Meaning and thought cannot be ascribed where such connections are not found.

This second component does not particularly concern the passage from observation and generalization to content ascription. It concerns and affects the whole process of interpretation and understanding of the other in all its moments.

A lot of things in an individual's actions and mental life must rationally cohere if he is to count as an agent, a thinker and a speaker. Accordingly, the interpreter must interpret so that the speaker turns out intelligible and rationally coherent in all such states and behaviors. The question of linguistic interpretation grows naturally into the question of general interpretation. As Davidson presents it, there is no way to find out about any of these things without learning about the rest. Beliefs and desires explain actions - linguistic acts, of course, included - and form the intentions that yield them[12]. One cannot identify the action without the knowledge of these mental states. In their turn, beliefs, desires and intentions cannot be discovered without figuring out the meanings of the person's utterances[13]. However,

---

[12] See, for instance, Davidson, 1963.

[13] See, for instance: "There is a principled, and not merely a practical, obstacle to verifying the existence of detailed, general and abstract beliefs and intentions, while being unable to tell what a speaker's words mean. We sense well enough the absurdity in trying to learn without asking him whether someone believes there is a largest prime, or whether he intends, by making certain noises, to get someone to stop smoking by that person's recognition that the noises were made with that intention. The absurdity lies not in the fact that it would be very hard to find out these things without language, but in the fact that we have no good idea how to set about

linguistic meanings only come about through linguistic acts[14], they are but abstractions from utterance meanings, and must be traced to the intentions with which words were uttered[15], and to their underlying beliefs and desires[16]. This is Davidson's picture and I am convinced he is mostly right.

All these states and actions must cohere in rational ways. In the rest of this section, I will try to detail some such ways. It is hard and somewhat artificial to explain and classify rational-coherence constraints apart from each other. Therefore, what follows is not meant as an attempt at a very sharp and exhaustive demarcation of matters, but rather more as a suggestive depiction of the diversity and complexity of the constraints involved.

First, and the most obvious, we find that the *logical relations* between contents determine, to some degree, the attitudes towards such contents that an interpreter is advised to attribute. "[W]e have no choice but to project our own logic on to the beliefs of another. In the context of the present theory, this means we take it as a constraint on possible interpretations of sentences held true that they are logically consistent with one another. Put another way, the policy is to assume the speaker's beliefs are logically consistent (up to a point at least)." (Davidson, 1980, 156-7) If the interpreter takes the speaker to believe that p, and to believe that not q, then he must not take her also to believe that $p \rightarrow q$.

Here I am using *logical* in the more strict sense of the word, and I mean those properties of contents that hold just in virtue of form or structure. "Logical consistency insures no more than the interpretation of the logical constants" (ibid.). What exactly fits into this class of constraints is, to some extent, an open question, with a somewhat contingent and convention-dependent answer: it is determined by what "we take to be the limits of logic and

---

authenticating the existence of such altitudes when communication is not possible." (Davidson, 1974a, 143-4) See also Davidson, 1973, 127, and Davidson, 1975, 162-3.

[14] See, for instance: "sentences are already a long way from most ordinary speech. We don't utter sentences, but rather tokens of sentences. Since communication depends on what we make of the tokens of others, and communication often succeeds, we can normally assume that others mean what we would mean if we uttered those sentences. This is something we can and do check up on, consciously or not, all the time." (Davidson, 2005b, 3). See also Davidson, 1993, 298.

[15] "an interpreter (correctly) interprets an utterance of a speaker only if he knows that the speaker intends the interpreter to assign certain truth conditions to his (the speaker's) utterance. ... my aim here is ... only to emphasize, following Grice, the central importance of intention in communication." (Davidson, 1992b, 111-2). See also Davidson, 1986, 93, Davidson, 2005b, 5, and Davidson, 1994b, 120.

[16] "... uttering words is an action, and so must draw for its 'teleological explanation on beliefs and desires. Interpretation is not irrelevant to the teleological explanation of speech, since to explain why someone said something we need to know, among other things, his own interpretation of what he said, that is, what he believes his words mean in the circumstances under which he speaks. Naturally this will involve some of his beliefs about how others will interpret his words." (Davidson, 1975, 161)

the list of logical constants" (ibid.).

"Further interpretation requires the assumption of further agreement between speaker and interpreter" (ibid., 157), and, indeed, there is more to coherence besides logical constraints. We can also recognize constraints regarding what we can describe as *conceptual relations* and *relations of rational support* between mental states. There is considerable overlap between both sorts of considerations, but I still find it illuminating to pause at the distinction.

> Beliefs are identified and described only within a dense pattern of beliefs. I can believe a cloud is passing before the sun, but only because I believe there is a sun, that clouds are made of water vapour, that water can exist in liquid and gaseous form; and so on, without end. No particular list of further beliefs is required to give substance to my belief that a cloud is passing before the sun; but some appropriate set of related beliefs must be there. If I suppose that you believe a cloud is passing before the sun, I suppose you have the right sort of pattern of beliefs to support that one belief, and these beliefs I assume you to have must, to do their supporting work, be enough like my beliefs to justify the description of your belief as a belief that a cloud is passing before the sun. (Davidson, 1977, 200)

The first question is that of providing enough *substance* to the states being ascribed. We seem not to be able to make sense of a thinker with just one thought - or just a few sparse thoughts. In particular, every thought seems to require a lot of beliefs accommodating it, enough to clarify the content involved. Contents of mental states are structured and complex wholes whose components enjoy a certain autonomy and generality. It is in part in virtue of the sharing of such components, let us call them *concepts*, that contents relate to other contents. To be interpretable as being in a particular mental state with some particular content, the agent must be recognized as a competent possessor and user of its constituent concepts. That implies being able to use them on a large enough range of occasions, as they feature in other contents and different states, in particular beliefs. Hence, the interpreter must already know or assume a lot about the agent's mental life before he is able to ascribe her any single thought or meaning.

The second question addresses the fact that mental states both implicate and are implicated by further mental states. Belief is the most flagrant case. In various ways and with different strengths, beliefs are explained and justified by further beliefs. However difficult it may be to render adequately explicit and clear the kind of considerations involved, the interpreter must be able to respond in the correct way to the presence and weight of these

complicated epistemic ties[17].

> ... it is hard to be precise about the rules for deciding where agreement most needs to be taken for granted. General principles are relatively simple to state: agreement on laws and regularities usually matters more than agreement on cases; agreement on what is openly and publicly observable is more to be favored than agreement on what is hidden, inferred, or ill observed; evidential relations should be preserved the more they verge on being constitutive of meaning.
>
> It is uncertain to what extent these principles can be made definite — it is the problem of rationalizing and codifying our epistemology. (Davidson, 1980, 157)

These relations of rational support, however, are not restricted to belief. Still in the intra-attitudinal holism[18] half, we find rational patterns also among conative attitudes. Take, for instance, the transitivity of preference[19]: if *A* prefers *b* to *a*, and *c* to *b*, she must also prefer *c* to *a*.

Lastly, there is also inter-attitudinal coherence to take into account. Principles of rationality bind together elements from all parts of the spectrum of propositional thought. A couple of examples are in order.

As already noted a few times so far, actions are understood and rationalized by desire-belief pairs.

> Giving the reason why an agent did something is often a matter of naming the pro attitude (a) or the related belief (b) or both; let me call this pair the *primary reason* why the agent performed the action. (Davidson, 1963, 4)

I get up and walk to the fridge because I want a beer and believe that there is beer in it. To be able to identify and understand an action, to see a movement of someone's body as a piece of intentional behavior, one must devise an adequate pair. Here, in Davidson's words, is the sort of rational coherence that can render adequate such a pair.

> A characteristic of teleological explanation not shared by explanation generally is the way in which it appeals to the concept of *reason.* The belief and desire that explain an action must be such that anyone who had that belief and desire would have a reason to act in that way. What's more, the descriptions we provide of desire and belief must, in teleological explanation, exhibit the rationality of

---

[17] I will have some more to say about the epistemological aspect of interpretation in Part II.a of this dissertation.
[18] Davidson explicitly invokes this distinction between intra and inter attitudinal holism in Davidson, 1995c, 13-6.
[19] See Davidson, 1974c, 237.

> the action in the light of the content of the belief and the object of the desire.
>
> The cogency of a teleological explanation rests, as remarked, on its ability to discover a coherent pattern in the behaviour of an agent. Coherence here includes the idea of rationality both in the sense that the action to be explained must be reasonable in the light of the assigned desires and beliefs, but also in the sense that the assigned desires and beliefs must fit with one another. (Davidson, 1975, 159)

As a second example of inter-attitudinal coherence, and again crucial to interpretation, we find the afore-mentioned links between linguistic acts and propositional attitudes. Speakers must hold a sentence true when they have a belief with the same content, they ask questions when they are uncertain, they formulate requests when they want something, and so on.

It would hardly make sense to try to exhaust all types of patterns of rational coherence between thoughts, but these collected examples are already richly representative of the kind of nexus that the interpreter must uncover and capitalize upon in his task.

Some of these cases are studied in Decision (or Rational Choice) Theory. Davidson's early training and work in the field[20] made him particularly attentive to such patterns, and he repeatedly appeals to that discipline in presenting and explaining the interpreter's methodology[21]. Often this is accompanied by the introduction of rigorous mathematical tools and methods, but they are inessential to my point and so I will not be exploring them in this dissertation.

As for some of the patterns exemplified, and others that the interpreter will certainly look for in the speaker, it may be difficult to ascertain whether they actually derive from *a priori* valid principles of rationality, or whether they are just very widespread conducts of thought and action. I do not think this doubt should bother us very much, I think it is fair enough to allow the interpreter to labor on such generalities, even if they are not clearly *a priori* sanctioned and necessitated.

We have seen now in some detail what pertains to each of the two components of the Principle of Charity that Davidson at times analyses separately. This must not overshadow the fact that there is a unity in the principle. I think Davidson grasps it well enough when he simply says that the principle of charity requires the interpreter to adopt a *"methodological*

---

[20] See "*An Interview with Donald Davidson",* Davidson and Lepore, 2004, 248-9.
[21] See, for instance, Davidson, 1974a, 145-8, and Davidson, 1980.

presumption of rationality" (Davidson, 1975, 159). Interpreters must look for the interpretation that reveals the other as rational as possible. However vague and minimal, this is the central point, the fundamental insight of which all other formulations are elaborations. The plastic, complex and indefinite nature of that central notion, rationality, suits our flickering understanding of what is comprehended in such a plan well.

1.2.5. A GENERAL CONCEPTION OF MEANING AND ACTUAL INTERPRETATION

A further and very pressing question that I have been postponing concerns the relation between what we have been considering, a fabricated method with which to tackle a very extraordinary interpretative setting, and the real nature of our interpretative practices. In this last section, I will defend that - and partially explain how - Davidson's views on Radical Interpretation imply and derive from a general conception of meaning, communication and language.

Davidson sometimes emphasizes, often in reply to his critics, that Radical Interpretation was never intended as an account of how people *actually* come to understand a natural language, but only about how they *could* come to achieve that.

> I have never claimed to know how children learn their first language. (...)
> I have never claimed to give an account of how field linguists arrive at their theories. (…)
> I am outlining what I claim could succeed, not what does. (Davidson, 1994a, 124-5)

This might suggest a certain limit to the consequences and reach of the thought experiment. What does a possible method have to do with actual interpretation? Davidson's answer to this question has been less than fully and clearly articulated.

It is undeniable, however, that he still expects his reflections on a possible method of interpretation to illuminate our real practices and the actual nature of language, meaning, and communication. From the beginning, along with disclaimers such as

> The second question, how we could come to have knowledge that would serve to yield interpretations, does not, of course, concern the actual history of language acquisition. It is thus a doubly hypothetical question: given a theory that would make interpretation possible, what evidence, plausibly available to a potential interpreter would support the theory to a reasonable degree? (Davidson, 1973, 125)

we also find clear expressions of Davidson's conviction that his inquiry concerns, in some sense, all linguistic communication - most notably, perhaps, when he declares that "[a]ll understanding of the speech of another involves radical interpretation" (Davidson, 1973, 125).

How can these two stances be reconciled? Part of the tension can be alleviated with an explanation provided in a footnote in "Radical Interpretation Interpreted".

> Reading over the papers I have written on interpretation over the last 27 years, it is clear to me that neither my terminology nor my views have held absolutely steady. I have sometimes used the term "radical interpretation" to refer to any interpretation from scratch, that is, without the aid of bilingual speakers or dictionaries, and sometimes to refer to the special enterprise of interpreting on the basis of a limited and specified data base. (Davidson, 1994a, 121, fn. 2)

When Davidson insists that his thoughts on radical interpretation are only about a possible method and not about the actual thing, plausibly he is just rejecting the claim that the increasingly complicated and artificial methods he coins in some of his texts, the *"special enterprises",* are faithfully enacted by real interpreters. Just as often, however, his considerations are meant as concerning and revealing the nature of real processes and competences - those effectively involved in *"any interpretation from scratch"* - though, of course, depicted at a very general and fundamental level.

How are we to tell what belongs to the *special methods* part and what is meant to grasp the real phenomenon? And what do these two topics have to do with each other? Davidson identifies some of the fundamental features present in all interpretation, and elaborates on that base with the aid of sophisticated formal methods - arguably adequate in theory, but implausible and unfit in practice - so as to reconstruct how an interpreter's competence could be built. He abstracts what he sees as the basics of interpretation and linguistic understanding, abandons what he deems inessential, and models both the target competence and the method of its acquisition in an optimally rational and rigorous fashion. Tarski's truth definition, the methods from decision theory, the limiting of the evidence available to correlations between sentences held true, or preferred true, and the circumstances accompanying the attitude make sense only in this context of rational reconstruction. However, underlying these special developments we find the fundamental insights that apply generally. In particular, taking into account the systematic relations between what is perceptibly the case and occasion sentences held true by the speaker, and the need to ascribe

meanings and thoughts to the speaker so as to render her rational, are basic rules that must, and do, shape all adequate interpretation.

As Davidson puts it, rather succinctly: "I believe that all successful interpretation depends upon the application of the principle of charity, and so, of course, that radical interpretation depends on it" (Davidson, 1994a, 122). The essential and pervasive role of the principle of charity, taken broadly - or, what amounts to the same, the need for rationalization - is the, or at least a, determining feature of interpretation in the Davidsonian picture I am arguing for.

And so it is fair to say, as anticipated above, that Davidson's investigations on radical interpretation amount to a foundational account of meaning. His ultimate point is not the speculation about an alternative and artificial method in itself, but to reveal the facts that effectively constitute meaning. Being ascribable, via rationalization, to some particular utterance of a particular speaker is, according to Davidson, constitutive of the utterance's meaning. As we will continue to explore in more detail in what follows, this does not mean that interpretation must always, or even frequently, involve the radical questioning and meticulous investigation described. But it does mean that all interpretation episodes are in some sense dependent, or grounded, on this most primitive type of inquiry.

The central role of charity and rationalization implies that the use of language is given a determined place in the bigger context of human interests and activities, one that provides it with both its viability and its purpose. Linguistic communication is depicted as a piece or aspect of a general enterprise of understanding rational agents, our fellow humans. It is the continuity between agents' linguistic manifestations and the rest of their human and intelligible life that allows them to use words in meaningful ways and be interpretable. It is the prospect of sharing an understanding of each other - not of each other's language - that sets interpretation and linguistic exchanges in motion.

In Davidson's words, "[a]ny attempt to understand verbal communication must view it in its natural setting as part of a larger enterprise" (Davidson, 1980, 151). This is the opening sentence of an important text where he lays down an ingenious method to explore the fundamental links between attitudes so as to find out simultaneously about the agents' meanings, beliefs and desires, starting only from evidence regarding what sentences an agent prefers true. The text is innovative in its technical details but the idea behind it is not new in Davidson, and is not limited to the particular service or elaboration delivered here. It is an

idea that he pursues and explores in several ways and directions throughout all his work, which says that the interpretation of the linguistic utterances of a speaker cannot be detached from the interpretation of the whole person or agent.

This interdependence between linguistic interpretation and general interpretation is true at those moments when the interpreter is still totally or mostly ignorant of anything that could count as the speaker's language – as has been, and will continue to be, the case in the situations depicted in Parts I and II.a of this dissertation – but also at those moments when a common language is practically shared – as I will defend in Part II.b.

We have had a good glimpse of the complexity involved in the interpretation of whole persons as we followed the disclosure above of several dimensions, and subdimensions, of the Principle of Charity. This was illustrative but by no means exhaustive. Rationalization is a rich, heterogeneous, and only vaguely delimited exercise. The broad rationality that the Principle of Charity imposes ranges widely. Interlocutors must share concepts, beliefs, logical norms, inferential patterns, values, interests, and more. A distinction such as the one presented above between Coherence and Correspondence introduces some nuance and order into the subject, but it must be viewed with some caution lest it mislead us. It would be a mistake to assume that the principle provides a clear and precise recipe for interpretation, or that it can be decomposed into more elementary ones that do. Instead, its application must ultimately be a job for our general and complete intelligence and judgment.

In the next chapters I will have still more to say about the diversity of mechanisms, abilities and reasonings - employed by the interpreter and ascribed to the agent being interpreted - that belong to the broad notion of rationality we are operating with.

Davidson's foundational theory has epistemological roots. The question that launches radical interpretation is about how to support and derive knowledge that suffices to yield interpretations of another speaker's words. Terms such as *knowledge, evidence, justification* and *verification* populate all these texts[22].

Davidson is working here with a particular type of epistemological exercise in mind. He is not considering indifferently any type of justification of linguistic knowledge or of meaning ascription - in particular, he is not employing an externalist conception of

[22] See, for instance, the first page of "Radical Interpretation", (Davidson, 1973, 125), and the preface to that collection of papers (Davidson, 1984d, xiv).

justification; he is not inquiring into the reliability of interpretation processes and mechanisms taking place outside the reach of the interpreter's critical judgment. It is only the rational, and potentially conscious, progression from ignorance to knowledge that interests Davidson. The question is pertinent because - and only because - to some degree, interpretation does indeed involve and require this type of rational learning. I will return to this claim, and to its defense, in the fourth chapter, where I will discuss the processes of lexicon acquisition by children learning their first language, and in the second part, where the interpreter's epistemology will be contrasted with Quine's naturalistic version.

It is no accident that a foundational theory of meaning is epistemological at its root. It is a consequence of giving a central place to the thesis that meanings are public. Davidson is not original in taking this stance.

> As Ludwig Wittgenstein, not to mention Dewey, G. H. Mead, Quine, and many others have insisted, language is intrinsically social. This does not entail that truth and meaning can be defined in terms of observable behavior, or that it is "nothing but" observable behavior; but it does imply that meaning is entirely determined by observable behavior, even readily observable behavior. That meanings are decipherable is not a matter of luck; public availability is a constitutive aspect of language. (Davidson, 1990d, 314)

If meanings are to be public and sharable they cannot be beyond the reach of an ordinary speaker's ability to interpret. Davidson restates this idea time and again, for instance, when he declares that "[w]hat a fully informed interpreter could learn about what a speaker means is all there is to learn" (Davidson, 1983, 148), or that "[t]he semantic features of language are public features. What no one can, in the nature of the case, figure out from the totality of the relevant evidence cannot be part of meaning" (Davidson, 1979, 235).

This has proven to be a very polemical position. Despite the strong intuitive appeal of the thesis that linguistic meanings must be public and sharable, the idea of conditioning the reality of something to the interpreter's ability to discover it has met fierce resistance. This is especially so since Davidson is willing to acknowledge quite significant deviation between the interpreter's verdict and the speaker's pre-theoretical take of the same phenomena. I will consider, in particular, the unpalatability of his position on meaning indeterminacy, and defend – in Chapter 3 - that there is a way of saving the core of Davidson's general interpretative approach while still rejecting inordinate consequences such as this. Very importantly, I will defend that being determined by radical interpretability - or by

interpretability through rationalization - is not incompatible with there being other factors concurring essentially to the determination of meanings. In Chapter 3, besides emphasizing Davidson's acknowledgement of speakers' and thinkers' first-person knowledge of their own thoughts and meanings, I defend the thesis that what allows this acknowledgement to cohere with Davidson's position on the constitutive interpretability of thoughts and meanings is the further requirement that both perspectives be concerted in a common appreciation of the semantic reality. Davidson presses for the need for such a concert in his recurrent appeals to the role of triangulation in thought, meaning, and interpretation.

In summary, then, this is the general and foundational conception of meaning, thoroughly Davidsonian, dictated by reflection on the most fundamental interpretative setting, that I am endorsing in this dissertation. Interpretation is an epistemic affair but, allied to adhesion to the thesis that meanings are public, reflection on the topic appears apt to yield metaphysical consequences, so that conditions on interpretation become conditions on meaning itself. Thus we can say that the meaning of a speaker's utterance on a particular occasion is determined by rationalization, that is, determined by her being interpretable, via rationalization, as meaning it. Linguistic interpretation via rationalization requires the interpretation of the whole person. It involves an adequate integration of the linguistic act so interpreted with the diverse remaining dimensions of that speaker's life and conduct. Finally, the adequacy of this integration is measured against a broad, complex, intuitive and shared standard of rationality that governs all these dimensions.

The public and sharable nature of meanings, the integration of linguistic practices in a wider human environment, and a very broad notion of rationality are the three key ingredients involved in the conception of meaning explored here. Having characterized it, it is now time to turn to its defense.

2. Justification of the Principle of Charity

The adequacy of the general conception of meaning and language distilled at the end of the last chapter and, in particular, the truth of the central thesis of this dissertation - that interpretation requires rationalization - require the validity of the principle of charity[23]. In this Chapter I turn to its defense.

Throughout the years, Davidson has explored several lines of justification of the principle. In a late text, a reply to a commentator in his *Living Philosophers* volume, he rounds them up very concisely in a bipartite argument.

> Briefly, the argument has two parts. The first part has to do with coherence. Thoughts with a propositional content have logical properties; they entail and are entailed by other thoughts. Our actual reasonings or fixed attitudes don't always reflect these logical relations. But since it is the logical relations of a thought that partly identify it as the thought it is, thoughts can't be totally incoherent, for if they were, they would be robbed of any possibility of being identified as one thought rather than another. The principle of charity expresses this by saying: unless there is some coherence in a mind, there are no thoughts, and if an interpreter is to grasp the thoughts of someone else, the interpreter must discover a sufficient degree of coherence to identify those thoughts.
>
> The second part of the argument has to do with the empirical content of perceptions, and of the observation sentences that express them. We learn how to apply our earliest observation sentences from others in the conspicuous (to us) presence of mutually sensed objects, events, and features of the world. It is this that anchors language and belief to the world, and guarantees that what we mean in using these sentences is usually true. (Davidson, 1999b, 343)

I will follow Davidson's lead here and conform my discussion of the topic to the division and categorization put forward in the passage, but I will also further analyze the ideas contained in these very terse enunciations, and distinguish in each case more than one attempted route to secure the validity of each principle.

In the first part - the arguments bearing chiefly on the coherence side of the principle -

[23] Note that the opposite does not immediately follow. It seems possible for the validity of the principle - i.e. the massive truth and rationality of agents - to be compatible with some alternative method of interpretation that does not actually exploit it.

I will consider appeals i. *to the nature of thought*, ii. to *semantic holism,* and iii. to *the holism of interpretation* in defending that thoughts can neither be discovered nor borne outside a sufficiently wide and coherent network of other states. I will also address a challenge to human rationality coming from a collection of studies in experimental psychology.

The second part will also be divided into three arguments in favor of correspondence, i. an *argument from exorbitance and inconsistency of justification standards*, ii. an *argument from interpretation,* and iii. an *argument from triangulation*.

2.1. ARGUING FOR THE PRINCIPLE OF COHERENCE

As already pointed out above, simple examples happen to be very convincing that thoughts must come in rationally organized clusters, and that single and autonomous thoughts extant outside a proper accommodating network of related propositional states seem inconceivable. There we considered the background of further beliefs needed to support a belief such as that "a cloud is passing before the sun" (Davidson, 1977, 200), but multiple other illustrations of this sort can be found scattered throughout Davidson's texts. In "Thought and Talk" he notes that "[i]f someone is glad that, or notices that, or remembers that, or knows that, the gun is loaded, then he must believe that the gun is loaded" (Davidson, 1975, 156-7), thus revealing how states other than beliefs depend on beliefs, and explaining immediately afterwards that "[e]ven to wonder whether the gun is loaded, or to speculate on the possibility that the gun is loaded, requires the belief, for example, that a gun is a weapon, that it is a more or less enduring physical object, and so on" (ibid.), thus elucidating how content of any particular state depends also on further beliefs. He makes very similar cases about what is required to have thoughts about trees (Davidson, 1982b, 98-9), black snakes (Davidson, 1995c, 10), man (Davidson, 2001e, 137), and more.

Indeed, it seems true that our intuitions speak in favor of some measure of gregariousness and rational articulation between thoughts. The point is quite generally conceded among critics and commentators[24]. Even Jerry Fodor and Ernest Lepore, who famously criticized *semantic holism* in their (Fodor & Lepore 1992), admit that there is an

[24] See Glüer (2011, 117), Ludwig (2004, 346-5) and Lepore and Ludwig (2005, 210-2)

instinctive inclination towards *thought holism*[25]:

> There is undeniably a pre-theoretic intuition that two people couldn't agree about only one thing. The intuition is that, if you and I agree that protons are very small, then there must be lots of other propositions we agree about too …. (Fodor & Lepore, 1992, 29)

even if they are *"not prepared to endorse this intuition straightaway"* (ibid). It also happens that these claims and points are not exclusive to Davidson. Stephen Stich, for one, uses the same sort of intuition pumping examples to argue for the holism of the mental.

> Shortly before her death, Mrs. T had lost all memory about what assassination is. She had even forgotten what death itself is. She could, however, regularly respond to the question, "What happened to McKinley?" by saying, "McKinley was assassinated." Did she, at that time, believe that McKinley was assassinated? For just about everyone to whom I have posed this question, the overwhelmingly clear intuitive answer is no. One simply cannot believe that McKinley was assassinated if one has no idea what an assassination is, nor any grasp of the difference between life and death. (Stich, 1983, 56)

The proliferation and cogency of illustrations notwithstanding, what exactly are the arguments here? It seems impossible to conceive a thought without a suitable mental system accommodating it, but why? Let us consider some attempts to formulate further reasons behind this intuitive impossibility. I believe we can distinguish three lines of explanation or argument behind the idea. The *argument from holism* actually fragments into three different forms: an *argument from the nature of thought*, an *argument from semantic holism*, and an a*rgument from interpretive holism*. I will take them one by one.

2.1.1. THE ARGUMENT FROM THE NATURE OF THOUGHT

The first proposal is that of defending that our holistic intuitions derive from, or reflect, the actual nature of our propositional attitudes. In some passages, Davidson appears satisfied with simply declaring that that is how our thoughts *intrinsically* are.

> The propositional attitudes provide an interesting criterion of rationality because they come only as a matched set. It may sound trivial to say that a rich pattern of beliefs, desires and intentions suffices for rationality; and it may seem far too stringent to make this a necessary condition. But in fact

[25] I will return to this distinction below.

> the stringency lies in the nature of the propositional attitudes, since to have one is to have a full complement. One belief demands many beliefs, and beliefs demand other basic attitudes such as intentions, desires . . . . the intrinsically holistic character of the propositional attitudes makes the distinction between having any and having none dramatic. (Davidson, 1982b, 96)

The idea would be that our intuitions are sound - and the interpreter's strategy and the general conception of meaning corresponding to it are adequate in this particular respect - because thought cannot be otherwise than holistic and coherent. Naturally, there is still something missing. It must still be explained how the intrinsic nature of something can have come to shape someone's intuitions and ideas about it. Davidson's response is to reverse the order of determination. It is not so much that the object of investigation informs our conception of it, but rather that our notions have a say in delimiting their objects.

> It should be emphasized that these maxims of interpretation are not mere pieces of useful or friendly advice; rather they are intended to externalize and formulate (no doubt very crudely) *essential aspects of the common concepts of thought, affect, reasoning and action*. What could not be arrived at by these methods is not thought, talk, or action. (Davidson, 1985, 92)

In this case, our notions of thought and agency are such that, among other constraints, they do not admit that single and independent mental states be ascribed, nor that too much incoherence prevails in the necessary accommodating system.

That some *Copernican revolution* is meant to be operative in Davidson's arguments is made fairly clear in numerous passages scattered throughout the years. In what Kathrin Glüer refers to as "a pretty confusing variety of pronouncements" (Glüer, 2006, 347), Davidson applies a number of heavy philosophical idioms to characterize the inquiry, arguments, and results he is engaged in when arguing for charity: "*conceptual*"[26], "*a priori*"[27], "*synthetic a priori*"[28], "*constitutive*", "*not a factual question*"[29] are employed fairly frequently, even if

[26] "I have been engaged in *a conceptual enterprise* aimed at revealing the dependencies among our basic propositional attitudes." (Davidson, 2005b, 73)
[27] "What makes interpretation possible, then, is the fact that we can dismiss *a priori* the chance of massive error." (Davidson, 1975, 168-9)
[28] "I suggest that the existence of lawlike statements in physical science depends upon the existence of *constitutive (or synthetic a priori)* laws like those of the measurement of length within the same conceptual domain. Just as we cannot intelligibly assign length to any object unless a comprehensive theory holds of objects of that sort, we cannot intelligibly attribute any propositional attitude to an agent except within the framework of a viable theory of his beliefs, desires intentions, and decisions. (Davidson, 1970, 221)
[29] "Smart asks 'whether people might not actually be approximately rational and consistent in their patterns of belief and desire'. In my view this *cannot be a factual* question: if a creature has propositional attitudes then that creature is approximately rational." (Davidson, 1985b, 245)

"*analytic"* or "*necessarily true"*[30] are, occasionally, scrupulously rejected.

Instead of trying to ascertain the precise nature or status of Davidson's reasonings and conclusions, I will simply endorse the largely shared[31] recognition that there is some cogency to his argument: our very general notions of thought, agency, meaning and the like, impose indeed some constraints as to what can and cannot count as instantiations of thought, meaning, and action. What in particular those constraints are, how strictly agents must conform to them, or how close they get to warranting the radical interpreter a possibility of success, are much more polemical issues.

Certain standards of rationality and certain patterns of articulation between propositional attitudes seem less negotiable than others. It is arguably harder to make sense of a violation of transitivity of preference or of a belief in contradictory propositions than, say, of failing to act in accordance with one's best reasons, or to form a belief that is best informed by the available evidence when that evidence is large and complex.

In other cases, it is the actual adequacy of some proposed rational structure that is disputed, not the degree to which thoughts or actions must conform to it. For instance, readers have been frequently averse to Davidson's idea that there is a conceptual dependency between thought and speech; those readers that reject this thesis will not be (as much) conditioned in their thought ascriptions by considerations regarding a creature's lack of speech. A different example of controversy - one particularly relevant to this dissertation and which will be discussed in its Part II.b - concerns whether or not some particular type of psychological background is required for a speaker to mean something with an utterance. Lepore and Ludwig address this possibility and acknowledge the Gricean analysis of meaning as a potential source of more rational structure.

> … there is potentially more to appeal to here than just general constraints on rational agency, for we are committed to taking subjects to be linguistic agents, agents whose repertoire includes speech

[30] "It may seem that I want to insist that decision theory (...) is necessarily true, or perhaps analytic, or that it states part of what we mean by saying someone prefers one alternative to another. But in fact I want to say none of these things, if only because I understand none of them. My point is skeptical, and relative." (Davidson, 1976a, 272)

[31] Among those commenting on Davidson, we see that Lepore and Ludwig indicate some sort of agreement. In Lepore and Ludwig, 2005, 210-3, they present these *"a priori* constraints dictated by the nature of the subject-matter, principles constitutive of what it is to be an agent or a speaker" (ibid., 219) and, unlike, for instance, their reaction to other "empirically derived constraints"(ibid., 220), they do not identify any problem in Davidson's appeal to them. Glüer (Glüer, 2011, 118-9) and Ludwig (Ludwig, 2004, 346-50) also express their endorsement of this idea.

> acts. If we can specify general constraints on patterns of attitudes that speakers must exhibit in performing speech acts in particular, then these patterns should be sought in subjects in any circumstances where it looks apt to treat them as performing a speech act. Thus, for example, we could bring to bear a theory that relates whatever intentions speakers have to the speech acts they perform, in something of the style of Grice's attempt to reduce utterance meaning to speaker's intentions (Lepore & Ludwig, 2005, 211)

A connection that is closer to hand between speech acts and mental states is, of course, the one most intensely explored by Davidson between belief and sentences held true.

To repeat, some degree of holism seems indeed imposed by our own notions, but too much is included under the title of coherence for a general and sweeping argument such as this to be effective in sanctioning all realizations defended by Davidson. The adequacy and plausibility of each hypothetical constraint, or set of constraints, must be assessed case by case, and it is likely that not all of them will turn out so clearly and irresistibly imposed by our notions.

2.1.2. THE ARGUMENT FROM SEMANTIC HOLISM

There is a potential alternative argument behind the claims and intuitions of holism and coherence. It is the second of the three listed above - the argument from semantic holism.

I must begin by clarifying the distinction already employed above between *semantic* holism (or *meaning* holism, or *content* holism) and *thought* holism (or *mental* holism, or *attitude* holism). The former can be said to correspond to any instantiation of the claim that the content of each contentful item from a certain totality - e.g. the words in one's language, the concepts in one's conceptual repertoire - depends on the content of most or all other items in that totality, and is (at least) partially constituted and defined by such dependencies and connections. The latter, thought holism, corresponds, as we have seen, to a different claim: the claim that propositional states can only take place in large enough and sufficiently rationally structured sets.

There are also frequent references - in Davidson and in the surrounding literature - to *belief* holism. Naturally, this refers to the dependency of beliefs on further beliefs, and is integral to thought holism.

Thought holism is the brand of holism that is most directly implied by the principle of

charity. Davidson endorses both types of holism, but here I am only concerned with thought holism. Thought holism and belief holism are weaker theses, in the sense that semantic holism implies them but is not implied by them. This asymmetry is explained by Peter Pagin in the following passage.

> In the case of belief holism, the stress is on the conditions for a person to have a belief with such and such a content. This is clearly a different question from that concerning how belief states depend on each other for having their content fixed. If there is a dependence, so that one belief state cannot have a particular content unless it is somehow connected to other belief states with appropriately related contents, then belief holistic claims follow. But the converse doesn't hold. There can be other reasons for belief holism than MH [meaning holism]. (Pagin, 2006a, 215-6)

He then goes on to illustrate how a belief can require other beliefs for reasons other than semantic holism. He proposes that "it is reasonable to claim that to have a belief that a gun is loaded the believer must minimally be able to distinguish guns from other things" (ibid.), and notes that this, together with two further equally reasonable claims, "first, that the only way of having that ability is having a grasp of functional features of guns (as distinct from perceptual features), and second, that grasp of functional features requires further beliefs" (ibid.), is enough to entail that any bearing of the initial belief is conditioned on the bearing of further beliefs. Pagin concludes that in a case such as this "[t]he need for having further beliefs is then epistemological rather than semantic: other beliefs are needed for some particular cognitive capacity" (ibid.).

Fodor and Lepore also reject that belief holism implies semantic holism - they develop and defend this rejection at length in the fourth chapter of (Fodor & Lepore, 1992) - but, as far as I could check, they offer no pronouncement on the inference in the opposite direction.

In certain passages, Davidson seems indeed to employ this second argument. Here he is endorsing a clear form of semantic holism:

> I think this is enough to ensure that some degree of holism goes with having concepts. Many concepts are fairly directly connected, through causality, with the world, but they would not be the concepts they are without their connections with other concepts, and without relations to other concepts, they would not be concepts. (Davidson, 2001e, 137)

Then, as explained by Pagin, the move from semantic holism to belief holism is dictated by the need to secure for concepts their constitutive connections to other concepts.

These connections are realized in the inferential links that hold between the beliefs in which the concepts figure, hence the idea that it is those links that determine content.

> A belief is identified by its location in a pattern of beliefs; it is this pattern that determines the subject matter of the belief, what the belief is about. (Davidson, 1975, 168)

This is the second argument. The crucial premise of course is that semantic holism is true. This, however, is a very polemical position, the object of intense and intricate discussion - especially during the nineties in response to Fodor and Lepore's critical assessment of the thesis (Fodor & Lepore, 1992). That being the case, and having myself nothing to contribute to the dispute, I have no option but to leave the argument at this stage as a mere open possibility, waiting for a more solid verdict on its fundamental premise.

2.1.3. THE ARGUMENT FROM INTERPRETATION HOLISM

Let us move on, then, to our third and final argument, the one from interpretation holism. It starts with the defense that interpretation requires the interpreter to be able to detect rational patterns and seeks, somehow, to infer from that a more general form of holism.

> We can appreciate why holism is not the disaster it has sometimes been portrayed as being if, instead of asking how the content of a concept or judgment is thought of by the creature that has the concept or judgment, we ask instead how an observer can size up the contents of the thoughts of another creature. (Davidson, 2001e, 137)

There is no way to interpret just one utterance or to ascribe just one thought. Each interpretation can only be supported in the context of a sufficiently systematic and coherent set of ascriptions. Without a background of further ascriptions, any interpretative hypothesis an interpreter manages to come up with will simply be wildly underdetermined. Consider, with Davidson, a very simple case which could, at first glance, seem to leave little room for doubts regarding the thoughts to be ascribed.

> Suppose I offer a person an apple and a pear. He points to the apple, and I record that he has chosen the apple. By describing his action in this way, I imply that he intended to point to the apple, and that by pointing he intended to indicate his choice. I also imply that he believed he was choosing an apple. (Davidson, 1974c, 237-8)

But then, as Davidson notes - in a different text, but referring to the exact same setting

- "a man who takes an apple rather than a pear when offered both may be expressing a preference for what is on his left rather than his right, what is red rather than yellow, what is seen first, or judged more expensive" (Davidson, 1975, 162-3). Only further interpretation can help us limit those options and bring about some determination. Only in the context of a much more complete grasp of a system of thoughts or meanings can the interpreter begin to clarify these fine-grained distinctions.

And then, of course, once the matter spills over from single thoughts into pluralities of thoughts, the already identified constraints of rational articulation jump in. The same rational requirements noted above, imposed by our notions of agency, thought and meaning, will be in force here determining what patterns of thoughts, actions and meanings we must look for in the other. Too massive or too blatant irrationality is not ascribable. What reasons could an interpreter possibly gather that would justify him in interpreting an agent as believing both *that mother is at home* and *that mother is not at home*, or as believing *that Paul is taller than John*, *that John is taller than George*, and *that George is taller than Paul*? What could convince an interpreter of something of the sort? Can he even make sense of such a possibility? Even with the benefit of a shared language, if the most obvious or literal interpretation delivers an absurdity, the natural reaction is for the interpreter to look for some misunderstanding along the way. In the case where there is not even such a strong presumption in favor of some standard meaning in a more radical setting, it seems that nothing at all would even be able to suggest irrationality or incoherence.

A particularly convincing way of defending the unacceptability of an ascription of too much irrationality and incoherence consists in noting how that would simply rob the interpreter of the ability to even identify the mental states in the person under interpretation, be they rational or not. For a particular propositional attitude to be ascribed, a background of other largely coherent attitudes must be in place. This holds with equal force for any eventual ascription of error, "the more things a believer is right about, the sharper his errors are" (Davidson, 1975, 168). Without such a background the identification of the attitude is endangered: "Too much mistake simply blurs the focus" (ibid.). Consider the following illustration of a mild (and entertaining) case of confusion and error,

> Today is a day of splendid triumph. Spain has a king; he has been found, and I am he. I discovered it today; all of a sudden it came upon me like a flash of lightning.
>
> I do not understand how I could imagine that I am a titular councilor. How could such a

foolish idea enter my head? It was fortunate that it occurred to no one to shut me up in an asylum. Now it is all clear, and as plain as a pikestaff. Formerly—I don't know why—everything seemed veiled in a kind of mist. That is, I believe, because people think that the human brain is in the head. Nothing of the sort; it is carried by the wind from the Caspian Sea.

For the first time I told Mawra who I am. When she learned that the king of Spain stood before her, she struck her hands together over her head, and nearly died of alarm. The stupid thing had never seen the king of Spain before! (Gogol, Memoirs of a Madman, 131)

There is a confused narrator expressing his insane realization that he is, after all, the king of Spain. There is no particle of truth to the fantasy. The idea seems to have emerged quite arbitrarily from a few disparate elements in the diarist's life – some news he read in the newspaper about a missing king of Spain, a certain discomfort with his own life, the ambition to improve his social position – and to have been equally poorly integrated in his experience of his ongoing affairs – see how he interprets his maid's reaction, how he feels lucky that no one had noticed that his previous state (according to him) was one of madness. Nonetheless, in a sense there still seem to be enough elements for something sufficiently close to the *belief that he is the King of Spain* to be ascribed to the narrator, at least for comical effect.

Things would become even harder for the standard interpretation were the narrator to add, for instance, *that to be a king is the loveliest of things since a king gets to be dressed in crystal and hang from a ceiling*, or *that Spain has always been his favorite spice*. If it went this way, an interpreter would soon be lost. The irrationality and error of the agent's beliefs gradually gives rise to a more fundamental difficulty, one affecting the actual identification of his beliefs, be they true or false, rational or irrational. Would it still be clear that he believed he was the king of Spain? And notice that the jamming of conceptual connections is here taking place at a fairly *material* level. The effects are even more drastic once we start disturbing more general and pervasive things like logical connections.

Commentators express ample agreement regarding the idea that interpretation requires coherence on the part of the interpreted agent.

That a certain amount of very basic rational thinking would be required for belief does seem quite plausible: If someone seems to violate basic logical laws all the time, for instance seems to permanently contradict himself, we very soon have no idea anymore what the person believes at all. And the same holds for someone whose beliefs do not seem to at all cohere in more 'material' ways. If

> we cannot at all make out why believing one thing would seem to speak in favor of believing another, it becomes unclear what it really is a person believes in the first place. (Glüer, 2011, 119)[32]

Again, however, just as with the argument from the nature of thought, limits of any sort remain uncertain and vague. Logical norms and certain principles governing choice and action will hardly admit violation. As for the rest, it is all very fluid and one struggles to formulate requirements of general application that are more precise and substantial than the simple claim that the interpreter must somehow be able to make sense of the agent.

As noted above, Davidson expects these insights into interpretation to reveal the actual nature of thought and meaning, thus consummating the move from interpretation holism to thought holism. Luckily, I do not have to defend or be concerned about the propriety of this inference here.

Thought is not my central concern here, but successful linguistic understanding is. I do not care whether or not a thinker is able to bear incoherent thoughts in the privacy of her own mind; I only care that even if that were possible she would not be able to share them with the rest of us. In this context, the truth of interpretation holism is just as good as it can get, that is, it exhausts all the support for the claim that interpretation requires rationalization that is extractable from an inquiry into the validity of the Principle of Coherence. The validity of the principle needs to be secured only up to the point where it confirms that no ascription of meaning can be such that reveals too much incoherence. If ascribed meanings and thoughts must be coherent and rational anyway, there can be nothing against a radical interpreter assuming it from the start in his approach.

2.1.4. HUMAN IRRATIONALITY

In the final decades of the last century, a number of psychologists, based on empirical results, brought human rationality into question. A number of (by now) famous experiments - developed, for the most part, in what has become known as the *heuristic and biases tradition* with classic references including, for instance, Johnson-Laird and Wason (1977), Nisbett and Ross (1980), and Tversky and Kahneman (1983)[33] - appear to reveal systematic errors in the

[32] See also Pagin, 2013, 242-3 and Ludwig, 2004, 350-1.
[33] See Samuels et al., 2004, for a detailed introduction to some of these experiments.

completion of tasks involving, for instance, logical, inductive and probabilistic reasoning. Since then, much has been written and discussed about how to interpret these and other similar experimental results, and whether, or to what extent, they recommend a revision of the traditional picture of man as a rational animal.

While some have used these results as sufficient reason to declare that human rationality is, in some important sense, or to some important measure, compromised, others have interpreted them more cautiously. Samuels *et al*. (2004) - which I follow closely in the next couple of paragraphs - collect, classify and elucidate a number of different reactions[34].

On one side of the dispute we find those who take the studies to reveal that humans actually reason in normatively defective ways, the relevant norms being those of formal disciplines such as logic, probability theory and decision theory. How apocalyptic a judgment on human rationality is issued by each of these authors still varies, depending on the extent of the flaws believed to have been identified, various theoretical options, and rhetorical ardor.[35]

On the side of those attempting to defend human rationality, different paths have been tried. Based on empirical studies of their own, *evolutionary psychologists* like Gerd Gigerenzer, Leda Cosmides and John Tooby argue that our reasoning competences are normatively sound but, due to their evolutionary history, have developed in such a way that they are more reliable in certain circumstances than in others. They try to show that in adequate conditions - for instance, when the information is presented in the right format - reasoners will perform in accordance with the relevant norms[36]. Other arguments maintain that the negative results are due not to defective reasoning competences but to the subjects' misunderstanding of the experimental tasks they are being asked to complete. The hypothesis is that certain pragmatic effects interfere systematically, and that this can explain much of the apparent negative performances[37]. A third form of resistance to the revisionary proposal questions the actual norms of rational reasoning, those allegedly violated by human reasoners in the experimental tasks. A number of authors have stressed various difficulties with the

[34] There are other critical surveys of the studies, positions and proposals on this question; for instance also Stein, 1996, Botterill & Carruthers, 1999, and Rysiew 2008.
[35] See Samuels et al. 2004, section 3.3. and Samuels et al. 2002, 236-242, and Rysiew, 2008, 1157-8, for some detail and references.
[36] See Samuels et al. 2004, sections 4 and 5.
[37] See Samuels et al. 2004, section 6.

operation of importing norms from formal theories to concrete and specific reasoning tasks[38]. Finally, it is also appropriate to add to this list Jonathan Cohen's early attack (1981) on the possibility of experimentally demonstrating human irrationality. Based on the idea that both particular responses to problems and the normative rational principles against which the first are measured are, ultimately, the finding, or intuition, of human intellects, Cohen argues against the claim that the experiments reveal faulty competences, and not merely faulty performances. His idea is that humans have to count as competent enough to reason in accordance with the standards or norms they have already managed to figure out. He appeals to the notion of *reflective equilibrium* to emphasize the intimate relation between norms and the particular judgments that should conform to them.

These questions and these competing views and responses are still the object of intense research and discussion. Here, however, is not the place for a detailed consideration, let alone evaluation, of the options on offer. Furthermore, as intriguing and auspicious as some of the proposals of the advocates of human rationality may be, we do not actually need them to defend Davidson's use of the Principle of Charity. Naturally, this is not to say that there is nothing to be gained from probing Davidson equipped with some of the findings and ideas that have briefly surfaced in these paragraphs. I am simply asserting that we do not need to go very far to develop a first, brief and convincing response to an urgent challenge.

Davidson never held or suggested that perfect rationality is to be assumed or expected. In fact, he explicitly asserted the opposite a number of times[39]. The identified errors taking place in those very specific, lab-construed experiments do not compromise the general rationality of the performers. Systematic and deeply ingrained as they may be, these errors are still confined to particular, and often marginal and specialized, types of reasoning tasks, and to particular aspects of the agent's rationality. There is much more to our traditional and commonsense notion of rationality - Ernest Sosa (1999), for one, calls attention to this fact.

What is more, the very identification of these particular mistakes can only be achieved in a context where the subject under test is assumed and confirmed as massively rational and correct in her understanding of, and coping with, the situation she is in and the task she is being asked to perform, not to mention in her wider understanding of the world, of herself and of the others around her. What emerges here is the more general Davidsonian theme

[38] See Samuels et al. 2004, sections 7 and 8.
[39] See, for instance, Davidson, 1982c, p. 99 and 1983, p. 150.

already identified in the previous section, namely, that the very possibility of error is dependent upon a vast groundwork of truth and correctness:

> This is not to deny the existence of irrational beliefs, actions, and emotions, needless to say. An action one has reasons to perform may be an action one has better reasons to avoid. A belief may be reasonable in the light of some but not the totality of one's other beliefs: and so on. The point is that the possibility of irrationality depends on a large degree of rationality. Irrationality is not mere lack of reason but a disease or perturbation of reason. (Davidson, 1982c, 99)

I conclude thus this very brief excursus into the famous and troubling empirical findings that began to surface some decades ago. I believe we can confidently maintain that they do not threaten the adequacy of Davidson's Principle of Charity[40], nor do they show that there is something wrong or unviable about a rationalizing approach to interpretation.

2.2. ARGUING FOR THE PRINCIPLE OF CORRESPONDENCE

The Correspondence component of the principle of charity is first introduced and put to use by Davidson in dealing with a very specific problem, that of suppressing the gap between what is open for a radical interpreter to see and the meaning and intentional ascriptions he aims to produce.

The manifest correlations between types of circumstances and utterances held true are notoriously insufficient to determine what is going on. They do not fully reveal what agents are thinking or talking about, whether they are right or wrong in their opinions, not even, strictly speaking, whether or not they are engaged in the type of linguistic practices that we are familiar with.

A brief excursus through other authors will provide us with some helpful contextualization and contribute to a fairer measuring of the difficulty. How silent and ambiguous can the observable facts be about meanings and contents of a fellow human is a recurrent theme in the recent philosophical tradition. One of its most pungent and influential elaborations is found in Wittgenstein's *Philosophical Investigations*, where, at a certain point the difficulty is summarized thus:

> This was our paradox: no course of action could be determined by a rule, because every

[40] Kirk Ludwig offers a similar response in Ludwig, 2004, pp. 348-50

> course of action can be brought into accord with the rule. The answer was: if every course of action can be brought into accord with the rule, then it can also be brought into conflict with it. And so there would be neither accord nor conflict here. (Wittgenstein, Philosophical Investigations, §201)

Any careful assessment of Wittgenstein's position on rule following is clearly beyond the scope of this dissertation, the subject being so very complex and polemical both in substantive and exegetical matters. However, at least a piece of the argument seems to have achieved general acceptance, namely, that no rule or intentional content is apt to be made fully manifest through any section of an agent's conduct non-intensionally described or apprehended. The same behavior, or the same set of behaviors, when non-intensionally described, conforms equally well to an infinite number of distinct rules.

Kripke, in his famous exploration of the topic, brought it even closer to our concerns. He focused on a more specific form of the general problem just presented: a non-intensional description of any speaker's personal history of use of any linguistic expression is logically compatible with an infinite number of different interpretations.

Kripke's central example involves the concept of addition and the linguistic expression 'plus'. A skeptic is invoked to question the grounds for any claim that someone is using the concept of addition and employing the word "plus" and the symbol "+" to mean *plus*, the regular mathematical function (Kripke, 1982, 8ff). Given our finitude, for each of us there is necessarily some number that is the largest he or she has ever computed. Assume - rather absurdly, but it makes no difference - that that number is 57. Now, the skeptic asks how one can be sure that in our previous uses we were not using "+" and "plus" to mean *quaddition* instead of *addition, quus,* instead of *plus* - a different function, identical to addition if the parcels are smaller than 57, but which, unlike addition, always delivers the same result, 5, if any of the parcels exceed that number.

He also, briefly, considers a concrete example, *table*. Again, the alternative is absurd.

> I think that I have learned the term 'table' in such a way that it will apply to indefinitely many future items. So I can apply the term to a new situation, say when I enter the Eiffel Tower for the first time and see a table at the base. Can I answer a sceptic who supposes that by 'table' in the past I meant *tabair,* where a 'tabair' is anything that is a table not found at the base of the Eiffel Tower, or a chair found there? (ibid., 19)

What facts can be referred to in order to prove that it was *addition* and *table* that was meant, and not *quaddition* and *tabair*? Kripke looks for possible grounds but finds none that

are proper. In fact, because his ultimate concern is ontological, not epistemological - whether there are semantic facts and what constitutes them, not just whether and how an interpreter can learn about them - Kripke looks for such grounds in places out of a common interpreter's reach. For instance, he considers appeals to the physical state of the body and brain of the agent described at some deep and complex level (ibid., 27ff), and to some possible "irreducible experience, with its own special quale, known directly to each of us by introspection" (ibid., 41). Here, however, I am only concerned with the interpreter's perspective.

What prompts me to include Kripke's treatment of the topic is not what he has to say about possible solutions, but rather the way he motivates the problem, especially his radical and awkward choice of possible alternatives to the regular meanings and concepts. When alternatives this absurd are conceded a genuine claim, it is because an agent's previous behavior is allowed to constrain the choice of possible meanings only in a very cautious and limited way. When this is the case, no increase of evidence can make a significant difference; an infinite number of different meaning ascriptions will still be equally satisfied by whatever sample one chooses. No doubt we will continue to find these alternative meanings rather strange and unnatural, but, considered from the same point of view that is willing to consider *quus* as a possible meaning, that will be *just our bias talking*. According to a certain austere conception of justification and impartiality, such empirical preferences and tendencies are not to be taken into account.

Quine is another illustrious case in the collection of thinkers who highlight the insufficiency of the evidence available to someone in the role of a radical interpreter. There are important and revealing differences between Quine and Davidson's projects and I will consider this contrast in detail in Part II.a of this dissertation. Here I just want to point to the fact that Quine's approach is equally caustic to any naive expectation that regular meaning ascriptions can be fully and safely grounded in independent evidence, thus dispensing the interpreter from contributing with biased assumptions of his own.

> … consider 'gavagai'. Who knows but what the objects to which this term applies are not rabbits after all, but mere stages, or brief temporal segments, of rabbits? In either event the stimulus situations that prompt assent to 'Gavagai' would be the same as for 'Rabbit'. Or perhaps the objects to which 'gavagai' applies are all and sundry undetached parts of rabbits; again the stimulus meaning would register no difference. (Quine, 1960, 46)

Nevertheless, Quine's examples of possible alternatives - at least the ones usually mentioned with respect to the 'gavagai' example: stages of rabbits, temporal segments of rabbits, undetached parts of rabbits, rabbithood and rabbit fusion (ibid., 47) - suggest a less radical take on the problem. For instance, there is, in Quine's examples, an implicit assumption of what we would perhaps like to describe as *continuity*. There is no *arbitrary* change, at some *arbitrary* point, from some *natural* meaning to another; no disruption, like in Kripke's, reminiscent of Nelson Goodman's predicates, *grue* and *bleen* (Goodman, 1954). But then again, any idea of continuity is relative to the predicates and classes one privileges, and so, perhaps, a fairer way to put it is simply to say that *we* do not experience Quine's alternatives as so unlikely, as so perverse, as Kripke's.

This sensible difference can, perhaps, be explained. The case is that Quine's emphasis is on a different difficulty. His concern is not so much the *infinitude* of the logically possible meaning alternatives that must be falsified – notice that he even allows infinite evidence. Quine's worry is about those alternatives that seem to resist any type of screening by means of unbiased observation alone. His point is that every observation or stimulation will be just as adequate or inadequate for affirming the presence of rabbits as for affirming the presence of rabbithood.

Underlying both Kripke's and Quine's problematizations we find an implicit and unclear standard of justification, one that is, in any case, impossible to meet. It is not a realistic demand. No interpreter can ever be in a position to exclude all possible meaning alternatives based on independent evidence before he is able to deliver an interpretation. But even if it is not clear how far a demand for justification can reasonably go, that does not change the fact that there is a vast gap between observation and content ascription that the interpreter must, somehow, overcome.

I will not be considering how Kripke and Quine work their way out of this problem. Here, I am only interested in the particularly vivid way in which they set it. Outside a proper context, one could fail to fully apprehend the immense power of the trivial assumption that our interlocutors share with us a common enough view of the world, at least regarding its most basic features. This Kripkean and Quinean background of skeptical questioning is meant to stress its real force and *silent* work.

The solution to be considered and evaluated here is, of course, Davidson's. It is his 'across the board' (Davidson, 1984d, xvii) application of the Principle of Charity, in

particular, the Principle of Correspondence.

Davidson draws what can be seen as a more realistic path across these same questions and concerns. If something is an agent, a thinker and a speaker, there are certain legitimate expectations regarding the possible contents of her actions, thoughts and utterances. As we saw in the previous chapter, Correspondence is not only about assuming the other to be largely right in her beliefs, but also that her beliefs and thoughts are about the same facts and features of the world. And so, regarding the logical gap between observations and ascriptions, what Correspondence proposes is that we should not attempt to regress too far, and to justify too much. Instead we must simply assume and enjoy the proper type of agreement that seems to hold among humans. As already quoted above, "the Principle of Correspondence prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances" (Davidson, 1991, 211). Yet, in other words, it urges the interpreter to "take what [the speaker is] caused to think and want as basically similar to what [the interpreter is] caused to think and want by the same objects" (Davidson, 1984a, 35-6).

The principle is quite intuitive. What Correspondence dictates is, undeniably, what we naturally tend to. What it declares must be the case is what we naturally assume to be so. Exhaustive justification is perhaps an impossible and misguided ideal but, be that as it may, it is at least certain that it has no place in our regular practices. We expect our fellow humans to be sensitive to, and interested in, much the same things as we are, to experience the world in much the same way we do, and we spontaneously produce our interpretations in conformity.

Even if, in practice, this ability to tune in to the same aspects of the world appears as sure and trivial as it can get, that should not deceive us about the presence there of an otherwise irresoluble epistemic problem. That problem, the bridging of the gap between observation and interpretation, is overcome only thanks to our mutual familiarity with the minds of our fellow rational beings. It returns the moment we decide to question or attempt to abandon or suspend our human and rational form.

And thus, of course, the Principle of Correspondence still admits questioning. What we naturally assume is not necessarily what we should. We might still want to ask: Is it really the case that fellow agents, thinkers and speakers agree in their experience of world? Can we trust interpreters' natural projections of their own thoughts and contents to render adequate and true interpretations? What reasons can we produce for the adequacy of appeals to

Correspondence?

In response, in what follows in this chapter, I will consider three different arguments supporting its adequacy. I will start by sketching an original argument with interesting but perhaps less than decisive results. Secondly, I will move on to a short, less ambitious but fully efficacious - for the purpose at hand - argument. Finally, I will introduce a third, more complex, ambitious and polemical argument, which I will not attempt to fully explore here, but that must still be signaled given its importance and assiduity in Davidson's later writings.

2.2.1. THE ARGUMENT FROM EXORBITANCE AND INCONSISTENCY OF JUSTIFICATION STANDARDS

I suggest that we start by fully conceding the impossibility of adequately responding to the skeptical challenge. There is just no way of solving the problem, i.e. of bridging the gap between the non-intensional and the intensional, in any way that lives up to the high epistemic standards implicit in Quine's and Kripke's problematizations. There is no way to prove to the skeptic that the speaker meant *rabbit* or *plus*.

With that firmly established, what we can do next is question the adequacy of the epistemic standards. Exorbitance already speaks against it. To opt for such a standard is to class all instances of interpretation as bad and unjustified, and to establish solid communication and understanding beyond human reach.

But there is more besides exorbitance. The skeptical puzzles in question also seem to require an inconsistent application of the high standard. The skeptic goes along with much less stringent conditions of justification in setting the stage for the puzzle to be able to arise. He seems committed to taking for granted that the agent being interpreted is a speaker and a thinker, that she shares with the interpreter an understanding of language and of linguistic practices, that she is interested and cooperative in the effort to achieve linguistic understanding. It is only after that, on top of that, that the skeptic emits his doubts about the content of her utterances and thoughts, only then do the justificatory scruples kick in.

This might not be a problem if the stage setting were simply accessory, if the initial concession of speech, thought and a cooperative disposition to the agent were a mere heuristic device to be also dispensed with once the skeptic had made his central point. But if the interpreter were really not to assume a communicative setting, all observations, all noted

contiguity between behavior and circumstances, would come to reveal nothing and determine, or underdetermine, nothing. The interpreter would no longer be interpreting, the question at stake would no longer be that of how to infer meanings and contents from the person's observable conduct.

If, on the other hand, one is serious about the person being a speaker and a thinker, it would be very difficult to justify the increased scruples the next moment. It is hard to accept that if the interpreter begins by knowing or assuming that the person in front of him is a thinker and a speaker, and has been, from situation to situation, repeating the same expression to him in order to let him know about its semantic property of denoting some portion of their shared environment, he can still genuinely bring himself to doubt whether she is talking about *plus, tables and rabbits,* or about *quus*, *tabairs*, *rabbit stages,* or even something utterly beyond his imagination and categories.

To attack the pertinence of radical justificatory challenges is to limit the scope for the inadequacy of the Principle of Correspondence in a very significant way. It seems that one cannot question Correspondence in such a critical and uncompromised way and still be seriously engaged in interpretation.

2.2.1. THE ARGUMENT FROM INTERPRETATION

The second argument for Correspondence is a straightforward version of the third argument considered above for Coherence: interpretation requires Correspondence, that is, there is no way to go about interpreting if not in conformity with the principle.

How could interpretation take place if speaker and interpreter were not tuned to the same aspects - especially the same perceptual aspects - of the world? How could an interpreter ascribe, to a speaker, thoughts about things and events he himself is not aware of, about features or aspects of the world that he himself is unable to detect and identify? Mutual interpreters, communicators, must, to a great extent, share their contents and concepts.

However, it is not enough that they share a subject; they must also concur on their opinions about it. This is particularly clear in a situation of interpretation from scratch, where there is no other way to start except by presuming the foreign speaker right. No clue could be taken from the other person's use of the language unless that use is supposed to be adequate. But this is also true when there is already good knowledge of each other's language or

idiolect. Occasional disagreement is possible, of course. But too much error, just like too much incoherence, cannot be made sense of. The consequence developed above for Coherence holds equally for Correspondence, that is, the interpreter would not even be able to identify any thoughts, be they true or false, correct or incorrect, if too much disagreement were admitted.

Vast and serious disagreement between interpreter and speaker is not really an open possibility. From the interpreter's inability to find the agent largely true and responsive to the same objects and events, what follows is simply failure of interpretation, not the suspicion of massive error, alien thoughts or ineffable meanings. I quote below yet one more passage from Davidson where he, once again, explains that agreement is a condition of understanding that it is impossible to agree to disagree too much.

> … everything depends on our ability to find common ground. Given enough common ground, we can understand and explain differences; we can criticize, compare, and persuade. The central point is that finding the common ground is not subsequent to understanding, but a condition of it. This fact may be hidden from us because we usually more or less understand someone's language before we talk with them. This promotes the impression that we can then, using our mutually understood language, discover whether we share their view of the world and their basic values. This is an illusion. If we understand their words, a common ground exists, a shared 'way of life'. A creature that cannot in principle be understood in terms of our own beliefs, values, and mode of communication is not a creature that may have thoughts radically different from our own: it is a creature without what we mean by 'thoughts'. (Davidson, 1984a, 37)

If we actually interpret each other successfully, if we actually communicate, the Principle of Correspondence must hold among us. This little seems to give us all we need from Correspondence in the present context. The thesis I am defending, that interpretation requires rationalization, requires Correspondence to be true only in those cases where successful interpretation actually takes place[41].

2.2.2. THE ARGUMENT FROM TRIANGULATION

There is a third argument for Correspondence with a very strong presence in

[41]For instance, it is irrelevant for the thesis whether or not we need to share our vision of the world with some hypothetical fellow thinker with whom we do not communicate. As was the case with the Principle of Coherence before, this is the most support the thesis can aspire to get from this quarter.

Davidson's later writings that demands a mention, but it is too complex and polemical for any adequate treatment to be attempted here. The discussion is rich and ongoing[42].

Unlike the previous one, this argument goes further than revealing the impossibility of interpretation without correspondence. It attempts to positively derive it from the processes and conditions of thought acquisition. Davidson argues that correspondence is a consequence of our contents being formed in a *triangular* setting, where both the world and our fellow humans *participate* in their determination. The critical part is being able to prove that this is so, not to mention offering any illuminating account of how this participation could be effected.

Davidson's stance here can be described as a peculiar form of *content externalism* in the sense that he takes the agent's *perceptual* and *social* environment to determine what she means and thinks.

First, it is the world itself, the things and events that populate it, that is taken to cause those same thoughts that have the world, with its things and events, as their object. And this causal relation must also somehow result as a content determining relation. As Davidson declares already in a relatively early text, and often reiterates afterwards, "we must, in the plainest and methodologically most basic cases, take the objects of a belief to be the causes of that belief" (Davidson, 1983, 151).

However, the agent's relation to the world, according to Davidson, is not enough to determine her thoughts. It is only the work of having them fit for communal use that concludes the job. According to Davidson, only a social setting, and the occasional discrepancy within it, can create the opportunity for concepts such as *error*, *belief*, *truth* and *objectivity* to emerge, and for the relevant cause of a perceptual belief to be clarified[43].

On the one hand, this is because "by yourself you can't tell the difference between the situations seeming the same and being the same" (Davidson, 1994b, 124), between following a rule and only thinking one is following the rule. And this, according to Davidson, is a difference that must have been grasped, and with it the mentioned concepts, by anyone who

[42] See Myers and Verheggen, 2016, Verheggen, 2013, Lepore & Ludwig, 2005, 404-13, and Gluer, 2011, 232-44.

[43] Still, this second agent must be close enough to the first in terms of her original endowment. The argument from triangulation is built on the premise that some standards of similarity are shared from the start, taken as an empirical fact. See, for instance, Davidson, 1999c, 165, and Davidson, 1992b, 118.

is to be counted as a thinker.

On the other hand, it is because, according to Davidson, only the presence of a second agent sharing an identical response to the same perceptual scene, together with the mutual awareness of their common reaction to the same object, can allow the disambiguation of the relevant cause from the multitude of causes concurring to the formation of a perceptual belief. The relevant cause, the one to be taken as the object of belief, is the "the nearest mutual cause of the joint reaction" (Davidson, 1999a, 41), that is, it is the closest item in the causal chain that is common to the two causal histories that end with the formation of the relevant belief in each agent.

There is something very plausible and attractive about the idea of one's thoughts being formed in a way that is responsive both to the world and to our interlocutors' solicitations. This would, to some degree, secure the type of correspondence and agreement we were seeking to prove. However, I am uncertain about the terms and success of Davidson's specific proposal, and in general about the prospects of a philosophical response to the skeptical challenges latent in these discussions. Accordingly, I abandon this line of reasoning and rest my case for the Principle of Correspondence on the previous arguments, especially the second one.

2.3. CONCLUSION

All in all, I conclude that we have enough of an a priori defense of the Principle of Charity in its dimensions that are most relevant to our purposes to sanction its application in the context of interpretation.

For the purpose in hand, the arguments from interpretation make the point in a clear and sufficiently conclusive way. By themselves, they do a fine job showing that interpreters must interpret in conformity to *Coherence* and *Correspondence* - or, at least, *Agreement*[44]. If the speaker is interpretable at all, her thoughts will be innumerous, cohere rationally, and picture the world in familiar ways. The remaining arguments, even if not decisive, reveal interesting avenues for further inquiry, and - with the possible exception of the argument from semantic holism - even add some form of support to the general idea.

[44] See the last paragraph of §1.2.4.1.

This conclusion supports the thesis I am defending in this dissertation. To claim that interpretation is guided by the Principle of Charity is the same as affirming that interpretation involves rationalization. In this chapter, we have established that it is at least licit to do so.

In the final chapter of this part I, I will complement the defense of Charity's role in interpretation with some empirical studies strongly supporting the idea that the principle is in fact employed in actual cases of interpretation from scratch. These studies show that word learning in the context of first language acquisition is best explained by attributing children the exercise of *charitable* reconstructions of their interlocutors' thoughts.

3. Two Challenges: First-Person Knowledge and Underdetermination

In this chapter, I address and resist two pressing objections to the possibility of successful radical interpretation of some unknown language, as Lepore and Ludwig (L&L, throughout this chapter) develop them in their pivotal commentary to Davidson's program (L&L, 2005). Both objections – the objection from *first-person knowledge* and the objection from *underdetermination* - target what L&L see as the radical interpreter's insurmountable disadvantage relative to the speaker regarding knowledge of the latter's meanings and mental states.

Addressing these two objections also gives me the opportunity to continue the exposition of my understanding of Davidson's ambitions, program and proposals.

In a first moment I will stress how the equilibrium and interdependence between perspectives – first- and third-person - over intentional and semantic facts is a key piece in Davidson's conception.

In a second moment I will turn to the conditions for radical interpretation brought about by this type of concert. Radical interpretation depicts a process or method through which an interpreter could arrive at a correct ascription of meaning and mental state to another speaker starting from a position of ignorance of her language. However, I will defend that the method is meant to work for speakers sharing a fundamental similarity, not for radical aliens. Between radical aliens no communication or understanding is possible. I will emphasize Davidson's idea that a vast agreement between interlocutors is a necessary condition for interpretation and understanding. Consequently, one should at once accept the impossibility of grounding meaning ascriptions in anything deeper and more solid than the interpreter's spontaneous familiarity with his interlocutor's mind.

3.1. INTRODUCING THE CHALLENGES

In their 2005 book, L&L present a certain assumption, which they claim to be central to Davidson's system, and which they try to prove wrong. This is "the assumption that the central concepts of the theory of interpretation - the concepts of meaning, belief, desire, intention, and the like - have their contents exhausted by their roles in accounting for the behavior of a speaker in a way that results in empirically equivalent theories of speakers, after all the evidence is in, that state the same facts." (L&L, 2012, 222)

They present "two challenges to this assumption which, if correct, would ensure that radical interpretation is not possible" (ibid.) and, ultimately, would undermine Davidson's general picture of language and interpretation[45].

"The first has to do with whether or not our having access to our own mental states independently of observing our behavior can be reconciled with Davidson's position on the central concepts of interpretation theory" (L&L, 2005, 222-3), and I will refer to this as "the challenge from first-person knowledge".

"The second challenge has to do with whether, from the interpreter's point of view, there are empirically equivalent but incompatible starting points for projecting interpretation theories"(ibid., 223). L&L argue that "[i]f there are, then the radical interpreter is not in a position to justify an interpretation theory for a speaker" (ibid.). I will refer to this as "the challenge from underdetermination"[46].

It is not clear whether to affirm the "central assumption of Davidson's approach to radical interpretation" (L&L, 2005, 222) is tantamount to affirming the radical interpretability of speakers - that will depend on how we interpret "exhausted" in the passage above[47]. What is clear is that the challenges, although targeting directly the first, the central assumption, present a threat to the second, radical interpretability, as well.

Furthermore, if what we have here is a two-pronged attack on radical interpretability, these challenges should also be expected to threaten the thesis being defended here, that interpretation requires rationalization, and the general conception of meaning and

[45] Both objections have a precedent and a subsequent history with other authors which I will not, for the most part, explicitly address.
[46] L&L return to these challenges in their comment on Gross's paper, L&L, 2012, 29.
[47] See Gross, 2012.

communication that support it and give it substance. Radical interpretation, in the broader understanding of the enterprise clarified at the end of Chapter 1, amounts to interpretation via rationalization - that is, a form of interpretation that targets both the speaker's thoughts and meaning and conforms to the Principle of Charity in all its richness and heterogeneity. According to the view I am defending here, this is the kind of process that underlies and supports, more or less directly, more or less explicitly, every instance of meaning ascription.

I find Davidson's proposal that *interpretation and rationalization reveal meanings because they constitute meanings* very appealing, and worth further exploration, even if not perfectly established as true, and I subscribe to it, however provisionally. But this is the line of reasoning that *the objection from first-person knowledge* is intended to interrupt. In affirming the existence of first-person knowledge, the first challenge intends to cancel Davidson's reason to be confident in the radical interpretability of thoughts and meaning, his conviction that they do not just happen to be radically interpretable but, in some sense, are actually determined by their radical interpretability. L&L appear convinced that there is no hope for the idea that meanings and thoughts are so constituted once an alternative access to semantic and mental facts is acknowledged that dispenses interpretation - first-person knowledge of one's own meanings and mind. L&L affirm that first-person knowledge of semantic facts attests their independence from interpretation and rationalization. In response, I will point out the compatibility between being accessible in non-interpretational ways and being essentially radically interpretable.

In their second challenge, L&L target the austerity of the interpreter's methodology - according, of course, to their own understanding of Davidson's proposal. The argument is supposed to show that the meager elements at the interpreter's disposal are not sufficient for successful interpretation. I believe, however, that they fail to do justice to the true richness and amplitude of scope of the Principle of Charity, as intended by Davidson. In response I will defend a different understanding of certain crucial features and moments of Davidson's radical interpretation experiment. By this alternative light, the enterprise will at once become more realistic and plausible, if less ambitious.

3.2. THE CHALLENGE FROM FIRST-PERSON KNOWLEDGE

Before addressing the first challenge, I start with a brief presentation of three aspects that L&L rightly identify as central to Davidson's position but which, according to my diagnosis, they then proceed to integrate in a distorted general picture of the original proposal[48]. It is this disputed rendition of Davidson's proposal that is being challenged by the first challenge. I will resist the objection by rejecting the interpretation.

3.2.1. THEORETICITY, ANTI-CARTESIANISM AND INDETERMINACY

The three aspects in question are Davidson's claims on the special status of interpretational[49] concepts, his anti-Cartesianism and his acceptance of underdetermination and indeterminacy of interpretation.

First, Davidson delivers frequent pronouncements on the special and *theoretical* nature of interpretational terms. He recommends some sort of caution when dealing and reflecting on issues of interpretation and intentional ascriptions, and questions the ontological autonomy of the objects of these concepts. He takes "everyday linguistic and semantic concepts [to be] part of an intuitive theory for organizing more primitive data" and warns that "only confusion can result from treating these concepts and their supposed objects as if they had a life of their own" (Davidson, 1974, 143). Later in this same text, he unveils a little more of what he has mind and draws an analogy between interpretation and its concepts and decision theory. He defends the view that "we should think of meanings and beliefs as interrelated constructs of a single theory just as we already view subjective values and probabilities as interrelated constructs of decision theory" (ibid., 146)[50]. Further analogies with concepts of other theories abound in Davidson's texts, but any other more direct characterization of the peculiar nature of interpretational concepts, or any explicit statement of the relevant similarity between the analogues proposed, eludes the reader.

Some uncertainty surrounding Davidson's position on the special nature of interpretational concepts notwithstanding, another thing is evident: Davidson's reticence about an independent semantic and mental reality converges with his rejection of a Cartesian

---

[48] I do not deny that occasionally we find passages suggesting L&L's reading. What I claim is that the reading I will be proposing fares better overall.

[49] By "interpretational concepts" I mean the concepts central to interpretation: meaning, belief, desire, intention, and so on.

[50] See also, for instance, Davidson, 1977, 224-5, and 1997, 74 ff.

picture of the mind and associated epistemology. According to the Cartesian picture and epistemology, propositional mental states and meanings belong to the category of things directly present to the mind, determined and complete in their spontaneous manifestation to the thinking subject, and, in particular, ontologically independent of any interpretation process. Other types of knowledge, knowledge of the world and of other minds, present an epistemological challenge, but not knowledge of one's own mind: introspection is secure, clear and simple. Interpretation, on the other hand, is seen as a murky and fallible business, and as something extraneous and subsequent to the minds themselves, something one exercises only to learn about the minds of others, minds whose character and reality are fixed in advance of any interpretative effort. Davidson determinedly rejects this idea of a direct and clear introspective access to mental and semantic facts operating in splendid autonomy from other forms of knowledge[51].

There is yet a last crucial strand to the story - *indeterminacy of interpretation*. According to Davidson, radical interpretation is likely to underdetermine[52] a system of meaning ascriptions - that is, as we saw in the first chapter, a truth theory for the language in question. Even after all the possible evidence is taken in, Davidson claims, it is possible that several distinct systems of interpretation can fit all the observable facts equally well. That is, it can happen that there is no way to decide, on the basis of all the available evidence, between two incompatible, or relevantly different, interpretations of the same utterance. According to Davidson, this would be due to the "flexible nature of the constraints".

> It is not likely, given the flexible nature of the constraints, that all acceptable theories will be identical. When all the evidence is in, there will remain, as Quine has emphasized, the trade-offs between the beliefs we attribute to a speaker and the interpretations we give his words. (Davidson, 1973, 139)

As interpretation is a holistic affair, tested only at some occasional moments of empirical friction, it should be possible to come up with different sets of ascriptions that are equally well confirmed by the data and further constraints. According to the particular method first delineated in "Radical Interpretation", the interpreter is set to work out two

---

[51] See, for instance, 1995c, 17.

[52] I will use "underdetermine" and "underdetermination" in a *neutral* sense, that is, I will use the expressions simply to mean that the facts in question are not enough to determine an option - a theory, an interpretation - independently of whether or not there are other facts that do. In this sense, *underdetermination of meaning by the relevant facts* is not opposed to indeterminacy of meaning but compatible with both determinacy and indeterminacy.

incognitas simultaneously, belief and meaning, in a way that is responsive to observed correlations of behaviors and types of circumstances, and the Principle of Charity. Several belief and meaning ascriptions seem possible at each isolated point, as long as there is a minimal coherence between the pair. The interpreter can conceive of the speaker talking about a rabbit, a tree or the weather, as long as he can conceive of her having the respective beliefs. The impractical freedom pertaining to each singular episode will be gradually reduced into something much more manageable by the necessity of integrating each ascription into a coherent system of meaning and thought. However, Davidson finds it plausible that in the ideal end - that is, when all the possible evidence has been taken into account, and each ascribable meaning or mental state is already accommodated in a fully determined system - there will still be enough looseness to allow for alternative interpretations. Underdetermination of interpretation is the claim that there is still some freedom at the level of complete systems, the claim that different truth-theories, entailing correspondingly different sets of thought ascriptions, can still accommodate, or re-describe, the totality of the speaker's dispositions for verbal behavior in such a way that she will be equally rational in each alternative.

This much would give us only underdetermination; there is still a crucial element missing to get to indeterminacy. It must also be that what is available to the radical interpreter exhausts all the semantic facts, that what is not determinable from the radical interpreter's position is not determined at all, that there would be no fact of the matter as to which of the hypothetical alternative systems of ascriptions is the right one. Underdetermination is an epistemic condition, but indeterminacy is ontological. To affirm the first is to affirm the impossibility of discovering, on the basis of the available evidence, which theory is the right one. To affirm the second is to affirm that there are no facts, knowable or not, that constitute the truth of one of the theories and the falsity of the others. Davidson's claim is that underdetermination detected from a radical interpreter's ideal position must reflect actual indeterminacy. As he puts it, "[w]hat no one can, in the nature of the case, figure out from the totality of the relevant evidence cannot be part of meaning" (Davidson, 1979, 235).

3.2.2. INTERPRETIVISM

L&L combine these elements to build what I take to be an inaccurate picture of Davidson as a radical *Interpretivist* (to give it a name). Interpretivism sustains that the special nature of interpretational concepts would consist of their being applicable only inferentially, from the interpreter's stance, to impose some artificial order in the domain of evidence. Ascriptions of thoughts and meanings to an agent - be it the interpreter himself or someone else - must allow the interpreter to organize, keep track, and make sense of that agent's behavior as externally observable. The belief that this is the only available criterion, that interpretational success responds to nothing more, allows Interpretivism to provide a tempting translation and explanation of *constitutive interpretability*: thoughts and meanings must be interpretable because that is all there is to them; being third-personally interpretable exhausts all their reality. This, L&L claim, is Davidson's view.

> Davidson treats the central concepts of the theory of interpretation of another speaker as theoretical concepts, whose function is to keep track of behavior. Viewed from this perspective, the role of a theory of interpretation is to identify and systematize patterns in the behavior of speakers in relation to their environment. If this is right, then we do not have access first to facts about speakers' meanings and attitudes, including our own. (L&L 2005, 11-2)

Interpretivism replaces first-person with third-person knowledge for the prominent role, taking the latter to be the fundamental form and origin of all others. This would be fundamental in sustaining the idea developed in the previous article of there being only one criterion of interpretational adequacy. L&L commit Davidson to this form of rejection of first-person, non-inferential knowledge of our own meanings and mental states[53,54].

This way, Interpretivism appears to create the sort of detachment that can be useful in making underdetermination and indeterminacy more palatable. If all that matters is to keep track of the agent's behavior, whatever exactly that means, maybe we can do it equally well in various different ways by telling systematically different stories and ascribing systematically different thoughts and meanings.

---

[53] See L&L, 2005, 227 and 243.

[54] I find it unclear whether they are defending that Davidson straightforwardly endorses this antagonistic position, or that he is simply committed to it because of his other claims, whether or not he knows it and is happy and conformed to it. In particular, L&L do not address the fact, to which we will turn in a moment, that Davidson explicitly acknowledges first-person knowledge and affirms it on a par, in terms of fundamentality, with third-person knowledge. In either case, my argumentation will include enough to oppose both claims: that Davidson actually professes it, and that Davidson is in any other way committed to it.

3.2.3. THE FIRST CHALLENGE TARGETS INTERPRETIVISM

It is against this radical *Interpretivism* that L&L present their first challenge, which we might, for commodity, reformulate thus:

C1. Davidson is committed to the rejection of first-person access to semantic and mental facts and this position is both false and incoherent with other aspects of his program.

L&L claim that, in rejecting the possibility of their application from a first-person perspective, Davidson's view of interpretational concepts is not only at odds with reality and our strong convictions on the subject - our belief in a direct and privileged form of access to our own mental states and meanings is deeply ingrained - but that it also generates internal inconsistencies within Davidson's system.

In one argument, they point to the fact that a radical interpreter must be able to make use of non-interpretational knowledge of his own experiences as a condition for engaging in any actual interpretation[55]. A second argument notes that for an interpreter to assume rationality on the speaker's part he cannot but assume that the speaker has first-person knowledge of her own states and of the meanings of her own words as well[56]. Since I do not intend to dispute any of these claims - that there is first-person knowledge, and that the radical interpreter needs to use it and to assume it in the agents he interprets - there is no point in lingering on the reasons and arguments behind them.

My response to L&L's *challenge from first-person knowledge* takes a different line. My plan is to oppose the Interpretivist interpretation of Davidson and offer, in its place, an alternative picture over which this objection has no hold. In the reading I propose, Davidson rejects Cartesianism but not first-person knowledge and its distinctive character. He affirms the theoreticity of interpretational concepts, but does not require them to be exclusively applicable inferentially from a third-person perspective. His position on underdetermination and indeterminacy does not derive from, or depend upon, a rejection of first-person non-inferential access to thoughts and meanings.

3.2.4. RESISTING THE FIRST CHALLENGE

[55] See L&L, 2005, 228
[56] See L&L, 2005, 365 ff.

L&L's reading, radical Interpretivism, might at first come as a natural interpretation of Davidson's position. However, there are sufficient indications that it cannot be exactly right. In particular, there is Davidson's flagrant acknowledgement of first-person knowledge and his rejection of any priority among the three forms of knowledge, or of any possibility of reduction of any form to any other, or others. He is particularly eloquent in this passage:

> I, like every other rational creature, have three kinds of knowledge: knowledge of the objective world...; knowledge of the minds of others; and knowledge of my own mind. None of these three sorts of knowledge is reducible to either of the other two, or to any other two in combination ...none is conceptually or temporally prior.... (Davidson, 1998, 87)

And this is not an isolated pronouncement, but something Davidson insisted upon quite frequently in his later writings. We find it again explicitly stated in a number of other passages[57], and more or less implicitly all over this period in Davidson's recurrent references to triangulation - first-person knowledge corresponds to one of the angles of the triangle defended by Davidson to be needed for thought and communication to take place. I will return to this issue in a moment.

Contrary to L&L's picture, Davidson does not affirm that one gets to one's own mental states and meanings only through interpretation and from a third-person perspective. Instead, he explicitly rejects this idea, for instance when he criticizes Ryle for having "stoutly maintained that we know our own minds in exactly the same way we know the minds of others, by observing what we say, do, and paint". He declares that "Ryle was wrong" and further explains that "[i]t is seldom the case that I need or appeal to evidence or observation in order to find out what I believe; normally I know what I think before I speak or act" (1987, 15).

Davidson does not fail to acknowledge the specific quality of first-person access, the "irreducible singularity of my direct acquaintance with the contents of my own mind" (1998, 91) - even if that does not mean indulgence on his part with any sort of naive conception of mental objects popping into one's consciousness.

Furthermore, he even notes the role of first-person knowledge in interpreting others. On a certain occasion, for instance, he points out the "fact so obvious it may escape notice: that the standards of rationality and reality on which I depend in understanding others are my

[57] See also, for instance, Davidson, 1991, 205-6.

own, and there can be no appeal beyond them" (ibid., 91). What is he here referring to if not to the fact that there is no alternative for the interpreter but to use his own thoughts and consult, first-personally, his own mind when trying to make sense of others?

The first part of the problem is done with. Davidson accepts first-person and first-person applicability of interpretational concepts. His position on both is not contrary to the natural and orthodox view. The question now is how to reconcile this with other aspects of Davidson's view, namely, with his anti-Cartesianism and with his position on the theoreticity of interpretational concepts - I will tackle the question of its compatibility with underdetermination and indeterminacy only afterwards, in response to the second challenge.

Davidson takes care to explain that this acknowledgement of first-person knowledge is not simply a relapse into Cartesianism. This is because these first-personally accessible thoughts and meanings "would not exist if it were not for a history of communication and experience" (ibid., 91). According to Davidson's proposal, first-personal experience must be acquired together with the two other forms of knowledge. The possibility of one's own private thoughts can only be secured in contact both with the world and other rational creatures. This interdependence of perspectives is profoundly anti-Cartesian. Still, I stress, it will not make one subjective experience any less first-personal. As Davidson puts it, "[t]hough we could not have been at the point of comparing notes without prior interaction, it is private notes that in the end get compared" (ibid., 91).

This also puts us on the track to an idea of interpretational concepts having a theoretical nature that is compatible with their also being first-personally, non-inferentially applicable. I propose that we take Davidson's assertion of their special nature to be referring to the fact that these concepts can only make sense, and be instantiated, as part of a larger, highly structured, and socially acquired theory. Below I will elaborate.

First, there is holism. It is not as if the thinking subject were simply the witness of full-fledged mental states showing up in her mind, each one by itself, complete, independent and irresistibly manifest in its specific identity - just like, say, an image on a screen, and then another, and then another. Something that we might want to call a "theory" must somehow come into place. Meanings and mental states, according to Davidson, are of a holistic nature. They can only be determined and supported in the context of a sufficiently vast and coherent system of other thoughts and meanings. Even in one's own case, the idea of an isolated thought or meaning, subsisting without any such background, simply does not make sense.

Their interrelatedness - logical, conceptual, inferential, justificatory - grants that each thought or meaning is both dependent and constitutive of the identity of others. The fact that an occurrence or identification of any thought or meaning is already committed to the existence of an adequately integrative background theory ought to count as one essential aspect of the theoreticity of interpretational concepts. This already gives us a good start in a plausible understanding, alternative to Interpretivism, of some of Davidson's pronouncements, such as his refusal to take "these concepts and their supposed objects as if they had a life of their own" (Davidson, 1974, 143)[58].

Second, there is the *concert of perspectives.* The requirement of agreement between perspectives, stressed three paragraphs above, with which Davidson forges his distinctive middle path between Cartesianism and Interpretivism, is also a decisive element in the conception of theoreticity I am trying to rescue from the rejection of inferentialism.

The concert of perspectives brings with it new substance to the notion of theoreticity. It means more structure and more theory, thus reinforcing the idea of thoughts' and meanings' dependency on a broad accommodating system. On top of the holistic constraints noted in the previous point, thoughts and meanings are further constrained by the necessity to agree across perspectives.

In the simplest and most basic cases, where, for instance, no deception is involved, this will mean that the thoughts a person thinks are the thoughts she would be ready to ascribe to her fellow human were she to believe that each other's situations match in the relevant sense. And vice versa, of course: a third-person ascription should roughly match a first-person ascription were the interpreter to find himself in the relevant same situation in which he pictures his interpretee. This is most dramatic in those paradigm cases of perceptual triangulation involving two agents and an observable object or event, where each participant compares her position and her response to the object or event with that of her interlocutor, to conclude, in the good cases, that they share a perceptual belief and that they are talking about the same thing. These are the basic cases, but the same dynamics are held to have a general implementation - consider, for instance, Davidson's observations regarding the need for

---

[58] The point can be made with different forms of commitment to holism. That is, one can still affirm the theoreticity of interpretational concepts while skeptical of the more polemical semantic holism - see §2.1 to §2.1.2 - as long as one remains convinced of thought holism. In any case, there will be enough structure involved in the entertaining of any thought and in the using of an expression to mean something, for a claim of theoreticity to be justified.

interpersonal agreement even in matters of value[59].

The concert of perspectives is what it takes for all three forms of knowledge and all three perspectives to be on a par, to be equally fundamental, as Davidson explicitly calls for in some of the passages indicated above. This is what I read in passages such as the following:

> … our propositional knowledge has its basis ... in the interpersonal. Thus, when we look at the natural world we share with others, we do not lose contact with ourselves, but rather acknowledge membership in a society of minds. If I did not know what others think, I would have no thoughts of my own and so would not know what I think. If I did not know what I think, I would lack the ability to gauge the thoughts of others. (Davidson, 1991, 219-20)

Notice that what we have here is a way to secure radical interpretability and the publicness of meanings alternative to that of Interpretivism. As indicated above, the interpretivist *strategy* is to abandon the source of possible discrepancy, first-person knowledge, allow only one form of access to the semantic and intentional reality, third-person, and thus ensure a perfect coincidence between being interpretable and being a meaning or a thought. We have now come upon an alternative route to the same effect: instead of extinguishing one of the perspectives, we can simply require first- and third-person to agree.

Why would this agreement be the case? First, for a long time, Davidson simply assumed the publicness of meaning. Gradually, however, he landed upon an explanation for this fact, a fact he started by merely accepting as *"obvious enough in itself"* (Davidson, 1990b, 62). As Davidson became more interested in the question of the emergence of thought and meaning, he came to consolidate a certain view of the process that, if correct, would account for meanings' and thoughts' constitutive interpretability.

Davidson holds that the ability to think, to mean and to ascribe meanings and mental states - in both cases, to ourselves and to others - is not a basic and uncultivated condition, but something acquired under propitious circumstances. The forming and maturation of the mind can only take place in a social milieu and in a causal relation to the objects of our thoughts. According to Davidson, the agent will not be capable of developing the ability to recognize her own states without at the same time developing the ability to think about the

[59] See, for instance, Davidson, 1991, 219-20, and 1995b.

world and about other people's mental states.

An alternative way of stating it would be to say that triangulation is required to determine meaning and thought[60]. That is, triangulation is more than a dramatization of the concert of perspectives. According to Davidson, it actually plays a role in their formation.

These questions have already emerged at the end of the previous chapter, §2.2.2. As I have said there, I find that there is something very plausible and attractive to the idea of one's thoughts being formed in a way that is responsive both to the world and to our interlocutors' solicitations, thus allowing for the right sort of accords - namely, knowledge and communication - to be effectively established. The claim itself - that, to some extent, thought and meaning require triangulation - is very tempting, but I remain unsure about the efficacy of the arguments that Davidson produced in its support.

I do not intend to investigate these arguments any further as their faults and merits are beside the immediate point. What I am trying to do here is, first of all, to reveal Davidson's ingenious balance between Cartesianism and Interpretivism - more specifically, that what he opposes to the primacy of first-person is not the primacy of the third, but rather a genuine concert of perspectives and of the three forms of knowledge - and, secondly, how this is still compatible with, indeed integral to, the theoreticity of interpretational concepts.

I take this to be enough as a response to the first challenge, the challenge from first-person knowledge. Davidson's proposal does not fall prey to it. First-person knowledge is explicitly accepted by Davidson and, so far, as we have seen, this acknowledgement is perfectly coherent with other aspects of Davidson's general proposal. We will continue this exploration in the second part of this chapter as we move to the topics of underdetermination and indeterminacy.

To demonstrate that the proposal is ultimately correct in all its elements is not the aim here. Naturally, I will leave some questions open, in particular, whether triangulation is required for thought and meaning formation. Furthermore, note that even if that were not the case - not always the case or never the case - this would still not show that there is no concert of perspectives. Nor would it show that thoughts and meanings are not radically interpretable, or even constitutively so. The hypothesis that triangulation plays a role in the formation of thoughts and meanings is but a possible explanation, and a possible reason, for holding those

[60] See, for instance, Davidson, 1982d, 105, 1992b, 120 ff, 1992a, 128 ff., and 2001b. See also Verheggen, 2013.

subsequent claims. Even if the hypothesis is proved wrong, those claims may still hold on different grounds[61].

To conclude the response to the first challenge, we seem to have here enough features - the holistic character of thought and meaning and the dependencies and concert between first- and third-person perspectives - to justify a claim of theoreticity[62], and the analogies drawn between interpretation and other theories, without having to give up on first-person knowledge. This also makes for a more plausible, and complex, understanding of interpretational concepts and their objects. Davidson is not denying the obvious - the peculiar and privileged access to one's own meanings and mental states. Instead, he is affirming something much more palatable, even sensible: that interpretational concepts and their application must, to some degree, be learned, and that this learning will go a long way in securing that those concepts be shared, and that there is accord between ascriptions produced from different perspectives.

3.3. THE SECOND CHALLENGE

It is time to move on to the second challenge developed by L&L. The threat here is that a convincing case for the underdetermination of meanings and thoughts from the radical interpreter's position could prove their radical interpretability false. Incidentally, underdetermination would also jeopardize the solution just defended for the first challenge. The hypothesis of agreement between perspectives - necessary, as we saw, to reconcile first-

[61] Besides, as already noted above, and as Claudine Verheggen (Myers & Verheggen, 2016, p. 67) also points out, before arriving at the argument from triangulation, Davidson would simply assume the publicness of meanings and thought, that is, their constitutive interpretability – see, for instance, Davidson, 1979, 235.

[62] Steven Gross argues that this idea of an agreement between perspectives is even suited to fit L&L's own explicit understanding of *theoreticity* - Gross, 2012, 230-32. L&L defend that, to qualify as "purely theoretical", concepts must have their content "exhausted by their application in the domain of evidence in a way that results in the content of the theories' theoretical claims not transcending their predictions about facts in the domain of evidence" (L&L, 2005, p. 225). Gross allows for two readings of "exhausts". In his prefered reading, the first-personal applicability of interpretational concepts is not a problem: "The interpreter, in applying interpretational concepts to himself, indeed does not apply them in systematizing his own behavior. ... But from this it does not follow that the concepts' content is not exhausted by its application to the relevant domain of evidence." (Gross, 2012, 231-2) What cannot happen, if interpretational concepts are to conserve their theoreticity, is for first- and third-person perspectives to issue incompatible judgements: "What does follow is that true ascriptions involving these concepts cannot transcend their predictions about facts in the domain of evidence. Thus, an accurate self-ascription involving interpretational concepts cannot conflict with warranted ascriptions that are made on the basis of all the evidence available to a radical interpreter subject to the constraints imposed in radical interpretation: the interpreter's self-ascription cannot conflict with the results of his being radically interpreted" (Gross, 2012, 232). On the same issue, see also Manning, 2012, 275.

person knowledge, anti-Cartesianism and the theoreticity of interpretational concepts - would be lost if interpreters could not arrive at thought and meaning ascriptions secure enough to match the speakers sure knowledge of their own thoughts and meanings.

We can reformulate the second challenge thus:

C2. The resources and constraints identified by Davidson are not enough to allow the radical interpreter to select just one justified interpretation. A number of alternative interpretations will be equally defensible from the interpreter's standpoint. This means failure, i.e. this means that radical interpretation, as proposed by Davidson, is not possible.

The hypothetical underdetermination would only be damaging if not accompanied by a correspondent indeterminacy affecting the agent's own perspective over the same meanings and thoughts. Davidson is quite ready to admit the likelihood of underdetermination precisely because he takes it simply to signal indeterminacy. As he sees it, the *fault* would not be epistemic but metaphysical. It is not the interpreter that falls short of reaching a justified option between several systems of interpretation that fit equally well all the available evidence. There is no fact of the matter determining which system of interpretation is the correct one.

L&L, along with a number of others, oppose this idea of indeterminacy, and they present their arguments[63]. I am not dealing with this issue here. I am not completely convinced by the arguments on offer on either side of the dispute and I am genuinely uncertain regarding the possibility of indeterminacy. Once more, having nothing relevant to add to a long and ongoing discussion, I have opted simply to drop the issue and leave it open as to whether or not some hypothetical underdetermination could be converted into indeterminacy and see its harmful potential neutralized. Instead, I have chosen to tackle the

---

[63] Their main argument focuses on a certain contradiction that should be patent from the interpreter's perspective. We can describe the radical interpreter's task as that of identifying pairs of sentences, one from the object language, the other from the metalanguage, with identical meanings. Alternative theories would relate the same object language sentence s with two different, non-synonymous, metalanguage sentences, p and q - being a competent metalanguage speaker, the interpreter could not fail to notice it. It would then be contradictory to accept both as true; the interpreter would end up affirming something close to: meaning of s = meaning of p, meaning of s = meaning of q, but meaning of $p \neq$ meaning of q. And so they conclude that it would be "incoherent for the interpreter to regard the different theories which he could confirm as both true" (Lepore and Ludwig, 2005, 239).

challenge from an alternative angle. I will defend the idea that it is far from certain that the radical interpreter would really be confronted with underdetermination.

L&L press for underdetermination by showing us how to build alternatives to the right ascriptions of meaning that are, they hold, equally justifiable from the radical interpreter's perspective. I will argue, however, that they are working with a rather impoverished understanding of the interpreter's perspective. This excessively austere view of the experience follows from an insufficient appreciation of the real richness and scope of the Principle of Charity.

However, this is a fault that Davidson himself sometimes appears to incur. I will defend the thesis that there is a tension between, on the one hand, Davidson's *'across the board'* (Davidson, 1984, xvii) application of the Principle of Charity and, on the other, some of the things he says about indeterminacy and, especially, the inscrutability of reference. I will start with Davidson and only afterwards return to Lepore and Ludwig's case for underdetermination. First, however, there are some points about radical interpretation that it will be useful to recapitulate at this point.

3.3.1. MORE THAN TRUTH, CHARITABLE INTERPRETATION

It is crucial to observe that radical interpretation is about securing more than a materially adequate truth theory for the speaker. The specifications of meaning it produces need to be such as to make the speaker more than just truthful, they ought to make her generally rational or intelligible. Extensional adequacy is insufficient to warrant that the theory's theorems specify the meanings of the object language sentences. As noted in the first section of the first chapter, this insufficiency was recognized by Davidson, and led him to add or stress further constraints on the theory. He imposes a holistic condition (1967, 26) - i.e. the theory should allow us to deal with all possible sentences of the language in a systematic way – and stresses the empirical grip afforded by the inclusion of demonstratives and indexicals (1967, 35), and he requires that the theorems hold nomological validity (1984b, xiv). All this could still not convince most people that the *interpretive* character of the theory was secure. Davidson's move towards Radical Interpretation is believed to be meant to help here, as it brought with it a crucial new constraint, the Principle of Charity, and the integration of the agent's linguistic career in the broader scheme of the speaker's life and environment,

familiarly human in its bare essentials.

L&L themselves emphasize the significance of the move. They explain that after having first attempted to defend that "a merely extensionally adequate truth theory for a natural language (i.e. one that is simply true) would thereby meet Tarski's Convention T or an analog for natural languages" (L&L, 2005, 75) - that is, would generate interpretative theorems - in a "second stage, when it became apparent that this condition was too weak, he appeals to confirmation by the procedures of a radical interpreter as an additional constraint" (ibid.). In view of this, it seems fair to assume that something will be missing if, in the end, we are not really getting anything new from this new setting. As it happens, I believe that we can only make sense of the proposed examples of, and robust confidence in, underdetermination of interpretation by, precisely, failing to acknowledge the additional resources bestowed upon the radical interpreter. My claim is that an adequate understanding of Charity drastically limits the plausibility of underdetermination.

3.3.2. DAVIDSON ON UNDERDETERMINATION

Davidson's most assertive and elaborate defense of underdetermination and indeterminacy of interpretation involves a specific form of indeterminacy, *the inscrutability of reference*[64]. That will be the focus of this section in which I aim to render patent the tension between Davidson's understanding of Charity and some of what he has to say about inscrutability, especially the examples he puts forward.

The thesis of the inscrutability of reference is supposed to arise from considerations regarding how every proposal for the logical analysis of sentences and the ascription of semantic values to their parts can only be indirectly confirmed or infirmed by how it fares at determining adequate results at the level of complete sentences. It affirms that, "there is no way to tell which way of connecting words with things is the right way; if one way works, there will be countless others that do as well" (Davidson, 1997a, 78). He goes on to explain:

> From a technical point of view, this means that for the standard satisfaction relation (...) we can substitute endless other relations without altering the truth conditions of any sentence or the logical relations among sentences. Since all the evidence for interpreting language must come at the sentential

[64] See, in particular, Davidson's 1979 "The Inscrutability of Reference", and his 1997 "Indeterminism and Antirealism".

> level … the result is that there can be no evidence that one of the satisfaction (or reference) relations is the right one. (ibid)

The idea, by itself, presents no immediate problem. It could well be that different reference and satisfaction schemes fitted equally well all the evidence *and other remaining constraints.* I am not here defending that underdetermination of interpretive hypotheses - and so, perhaps, indeterminacy and inscrutability - is obviously incompatible with Charity. What I do want to stress, though, is that to fabricate an alternative interpretive system around such a vast and powerful constraint would be much harder than is usually conceded. With an adequate understanding of the principle, we can easily discard as inept the kind of examples of alternative possible interpretations put forward by Davidson and others. Here is one such example:

> Suppose satisfaction relation *s* maps the word 'Rome' onto Rome, and the predicate 'is a city in Italy' onto cities in Italy. Then the truth definition will show that the sentence 'Rome is a city in Italy' is true if and only if Rome is a city in Italy. Now consider another satisfaction relation *s'* which maps the word 'Rome' onto an area 100 miles to the south of Rome, and the predicate 'is a city in Italy' onto areas 100 miles south of cities in Italy. The truth definition will now say that the sentence 'Rome is a city in Italy' is true if and only if the area 100 miles south of Rome is an area 100 miles south of a city in Italy. The truth conditions are clearly equivalent. The thesis of the inscrutability of reference contends that there can be no evidence that *s* is any better than *s'* for interpreting the sentence 'Rome is a city in Italy'. There is no telling what a sentence is 'about', or what someone is thinking about. (ibid.)

And now we must ask: how is this second interpretive hypothesis not in flagrant violation of Charity? It is not indifferent for the speaker's overall rationality whether she is generally talking and thinking about the things she has in front of her, she points to, she visits or lives in, or whether she is generally talking and thinking about something taking place *"100 miles to the south of"* those things and places.

Davidson also invests great importance in the idea that "causality plays an indispensable role in determining the content of what we say and believe" (Davidson, 1983, 150), and that "we must, in the plainest and methodologically most basic cases, take the objects of a belief to be the causes of that belief" (ibid., 151). And so, at least in those basic cases of perception sentences, this should be enough to free the interpreter from the need to consider alternative interpretive hypotheses that ascribe to the speaker pronouncements about objects too remote - e.g. 100 miles to the south of the perceptual scene - to play any credible role in the causation of the belief and respective utterance. This matters to more than a small

or marginal portion of the theory, since most terms have good chances of appearing in perception sentences - "Rome", of course, included. Furthermore, Davidson requires that, in these basic cases, interlocutors ascribe meanings and sentences so as to identify a shared cause and object for their linguistic interchanges. He affirms that "[c]ommunication begins where causes converge: your utterance means what mine does if belief in its truth is systematically caused by the same events and objects" (ibid., 151). Accordingly, only an interpreter perceptually interested and moved by something 100 miles away would seem to be in an adequate position to ascribe a similar state to his interlocutor.

One could maybe try to argue that examples such as this are supposed to have a merely illustrative character and function, that they are not meant as actually valid. In "The Inscrutability of Reference", Davidson is explicit about it, and he acknowledges that the alternative explored there – things' shadows, instead of things themselves - does "not exactly fill our bill" (Davidson, 1979, 230). About the possibility of better examples, Davidson declares simply that he "assume[s] that there are permutations of the requisite kind that demand no fiction" (ibid.).

This idea could help to avoid an otherwise mystifying inconsistency on Davidson's part. But, even if the point were conceded, it must still be noted that Davidson's strong confidence in the existence of adequate alternatives is expressed with no reservation, and that such a position, just by itself, is yet too hasty and lacking adequate support.

In the end, however, the most plausible account is perhaps that both Davidson's confidence in the inscrutability thesis, and his lack of rigor in the picking of an example of alternative, are based on his correct assumption that the existence of infinitely many unnatural *proxy functions*[65], allowing for the derivation of infinitely many *materially adequate* truth-theories from any materially adequate original truth theory, is a certain thing. But if this is all there is to Davidson's defense and arguments, it betrays exactly the gross neglect of the Principle of Charity that we were seeing if we could absolve him from. His exclusive concern seems to be the preservation of truth across alternative interpretations, and there is no thought for the general intelligibility and common humanity of the speakers.

There is a flagrant tension, and I cannot find any way to alleviate it while saving both

[65] A proxy function determines "some one-to-one mapping of every object on to another" (Davidson, 1979, 229). Davidson explains the idea in more detail, for instance, in Davidson, 1979, 229.

Charity and Davidson's apparent positions on inscrutability. The endorsement of a strong Principle of Charity is the more fundamental position here. It is a crucial tenet of Davidson's program, one that he simply cannot surrender. Accordingly, I cannot avoid concluding that there must be something wrong, or at least misleadingly formulated, with Davidson's appreciation of the inscrutability hypothesis.

3.3.3. LEPORE AND LUDWIG ON UNDERDETERMINATION

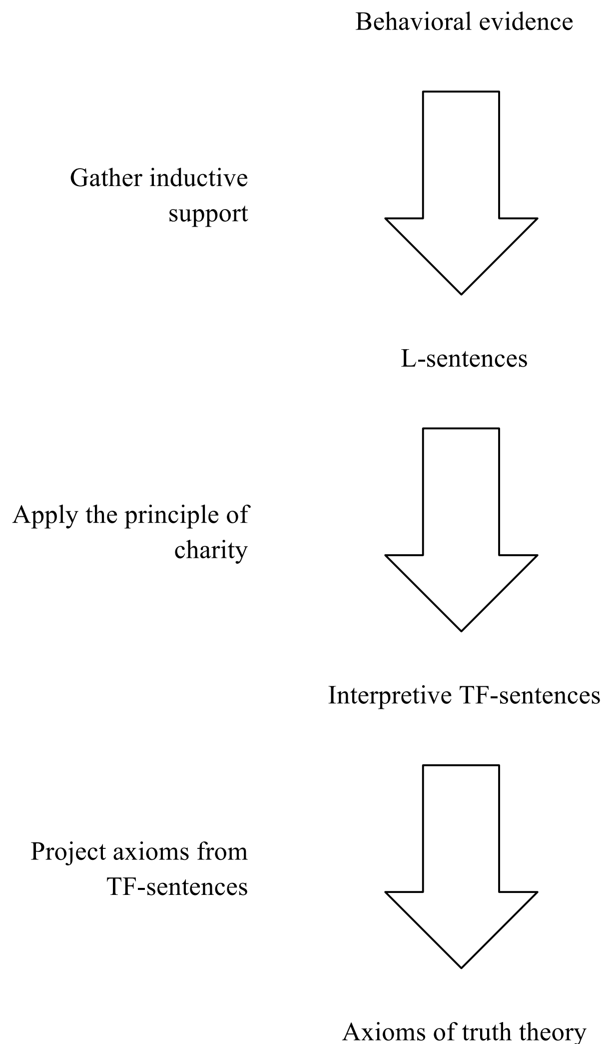
Lepore and Ludwig share my concerns regarding Davidson's position on inscrutability. They note that in Davidson's case for inscrutability it appears that:

> … nothing restricts what object a permutation can take the originally demonstrated object to. The object may be one with which the speaker has no causal or perceptual contact, and no way of knowing about, from the point of view of the interpreter. It is hard to see how assigning such an object to a use of a demonstrative could be compatible with the principle of charity, or any common-sense principles of interpretation. (L&L, 2005, 379)

However, despite their good judgment in this case against Davidson, at certain moments in their book, they display a comparable inattention to the Principle of Charity. This is noticeable, namely behind their attempt to prove underdetermination of interpretation from the stance of the radical interpreter.

What L&L try to prove is that the same evidence will always support equally well sufficiently different, or "*non-equivalent*", generalizations. The latter will in turn issue problematically different T-theories. Let us consider this in some detail.

They start by referring the reader to a scheme (ibid., 229, fig.1) representing the different stages of the process the radical interpreter must go through, which I reproduce here for perspicuity.



The first step in the process takes the interpreter from hold true attitudes to *L-sentences*. "L-sentences[66] are generalizations about the conditions under which speakers hold true sentences" (ibid.), more precisely, they "are universally quantified inductively supported generalizations, which employ no more concepts than those already at play in the evidence for them" (ibid.). They "provide the evidence for the next stage in the process" (ibid.).

At the end of this stage, the interpreter is supposed to have collected a good number of such generalizations upon which he is to try to construe the interpretative truth theory generating ascriptions of meaning, or truth-conditions, to the utterances of the agent being

[66] I called them *generalizations* or *G-sentences* in Chapter 1 of this dissertation.

interpreted. The collection of such generalizations will form the *"data set"* that is to support that theory.

Next, L&L explain how what comes next - TF-sentences[67], and the theory itself, with its axioms and theorems - will simply follow and perpetuate the option blindly exercised at the moment when the interpreter first described what he was seeing in the way that most spontaneously occurred to him. Imagine that the interpreter had naturally assumed that the speaker's utterance of "Le chat est sur le tapis" had something to do with *cats* and not with, say, *temporal stages of cats*. This assumption is then transported to the generalization, and then, ultimately, to the theory itself.

Finally, the trouble L&L point to is that spontaneity and naturality notwithstanding, the same observations, the same behavioral evidence, can equally well support different, "*non-equivalent*" data sets[68], and, consequently, different, "*non-equivalent*" theories[69] and interpretations.

By now we have returned several times to this general problem of the impossibility of solidly grounding content ascriptions in manifest behavior, but we must articulate some of the details of L&L's specific take on it so as to be able to try to meet their particular challenge afterwards.

They present as a sufficient condition for the existence of "non-equivalent data sets confirmable from the interpreter's evidence"(ibid., 231) that "there be two law-like sentences for an occasion sentence s of form (L) expressing different laws that can appear in data sets" (ibid.). The following scheme fixes "form (L)":

(L) Ceteris paribus: S holds true s at t iff p. (ibid.)

Then they state that to prove that the sufficient condition is fulfilled, one needs just to come up with "true counterfactual supporting[70] biconditionals of form (EQ), with identical

---

[67] I called them *proto-theorems*, or T-sentences, in Part I

[68] Data sets whose correspondent L-sentences right hand sides do not translate one another (Lepore and Ludwig, 2005, 229).

[69] Truth theories whose correspondent T-sentences right hand sides do not translate one another (Lepore and Ludwig, 2005, 229).

[70] Lepore and Ludwig carefully follow Davidson's indication that the theorems should have the value of rules, and hence support counterfactuals. However, they qualify their statements "with *ceteris paribus* ('all other things being equal'), because these laws are rough laws which can be assumed to hold only relative to certain conditions obtaining" (L&L, 2005, 163).

ceteris paribus conditions, where what replaces ‘p’ and ‘q’ in (L) and (EQ) are nonsynonymous” (ibid.232). Here is the scheme that fixes “form (EQ)”:

(EQ) *Ceteris paribus*: *p* iff *q*

Finally they give us five recipes for the production of (EQ) equivalences. Once more, my response will be that when all constraints on interpretation are really taken into consideration - in particular, the Principle of Charity - the results are not so obvious as to whether any of the alternative data sets can actually be accepted by the radical interpreter. I will briefly run through those recipes before referring them to a plausible antidote in Davidson’s stock.

First, however, there is an obstacle - not a genuine one, I would say, but something of a hermeneutical genesis - that must be cleared away. The way L&L present the case, having located the root of the underdetermination at the level of L-sentences, one is led to assume that the problem lies outside the Principle of Charity’s influence. But it is not so. The Principle of Charity must govern the whole process, including the choice and description of evidence and the inductive generalizations. One cannot imagine it to be otherwise, especially once it has become clear how the choice and description of the interpreter’s evidence matters crucially for the final result. If the Principle of Charity is to be of any relevance in the process, it cannot be absent from those first stages. In omitting any intervention of the Principle in those first stages of interpretation, the scheme and division presented above are not only artificial but also somewhat inadequate and misleading.

Let us now turn to L&L’s recipes for the production of (EQ) equivalences, and the consequent alternative generalizations, or L-sentences.

Recipe a) tells the interpreter to “form a conjunction of ‘p’ and a sentence which expresses a nomically necessary truth” (ibid., 232). For instance, instead of going for something like:

L_1: ceteris paribus: S holds true s iff *it has started to rain near S.*

the interpreter could just as adequately go for

L^a_1: ceteris paribus: S holds true s iff *it has started to rain near S and 2+2=4.*

Recipe b) tells the interpreter to "substitute for a predicate in 'p' a nomically co-extensive nonsynonymous predicate" (ibid.). In this case the idea is that the interpreter can freely choose, for instance, between chemical descriptions - *"H2O"* , *"element with atomic number 79"* - and phenomenal descriptions - *"watery stuff"* and *"golden element"* - of the same objects.

Recipe c) tells the interpreter to "substitute for 'p' any sentence nomically equivalent (but nonsynonymous)" (ibid.). They elaborate an interesting example. They affirm that "if a distal event expressed by 'p' suffices for a speaker to hold true a certain sentence, it will also suffice for the truth of sentences about intermediate conditions necessary for the transmission of the causal efficacy of the distal event to the speaker" (ibid.). That is, instead, for instance, of a sentence describing rabbits, the interpreter can go for a sentence about "a certain kind of pattern of irradiation of one or the other or both of the speaker's retinas" (ibid.).

Recipes d) and e) are slightly more complicated as they involve some interference with the ceteris paribus qualification. Recipe d) tells the interpreter to move p into the background conditions in exchange for some arbitrary condition that is found there. The ceteris paribus clause comprehends a vast and unspecified number of conditions. It was introduced by L&L to accommodate the fact that things sometimes go wrong, and for all sorts of motives: a*ll things being equal*, the speaker will hold the sentence true, but not if she is making some sort of mistake, is in shock, or is having a stroke. Note that each of those background conditions actually contribute just as decisively as the original p to the holding true of the sentence by the speaker.

Now the idea is to take any one of those conditions usually in the background - e.g. *S is not having a stroke* - and move it into prominence. Let us use "*Ceteris paribus^{d}*" to represent all the usual background condition minus the chosen one. At the same time, the original p takes the inverse trip and recedes into the background.

Returning to the above example, with d) Lepore and Ludwig are defending that the interpreter is free to choose between L_1 and L^{d}_{1},

L_1. ceteris paribus: S holds true s iff it has started to rain near S.

L^{d}_{1}. ceteris paribus^{d} & it has started to rain near S: S holds true s iff S is not having a stroke.

Recipe e), the last one, is hard to distinguish from d). Lepore and Ludwig, again referring to the ceteris paribus condition, affirm that, “[i]t is overwhelmingly plausible that if there is one way to conditionalize on causal conditions to secure the nomic necessity of a biconditional of the right form, then there are many” (ibid., 233). And they develop the idea, explaining that “for example, if there are conditions in the distal environment relative to which a speaker holds true a sentence s iff p, there will be more proximal conditions relative to which the speaker will hold true s iff q, where that p and that q are non-equivalent conditions” (ibid.).

If I understand them correctly, proposal e) is very close to d). It tells the interpreter that he can substitute *p* for a nonsynonymous *q* as long as he compensates for it with some corresponding change in the specification of the background conditions.

3.3.3.1. RESISTING LEPORE AND LUDWIG’S CASE FOR UNDERDETERMINATION

These are the five formulas presented by L&L for the generation of alternative data sets that confirm, they defend, the underdetermination of meaning from the radical interpreter’s perspective. While it is certainly true, in every one of these five cases, that the alternative generalizations would preserve the truth, and even the nomological truth, of an originally nomologically valid L-sentence, it is also true, I defend, that the new L-sentences would no longer be convertible into statements of truth-conditions, i.e. T-sentences or TF-sentences. My response to all five alternatives is basically the same, and a very thin one: they are too odd and the interpreter would not be able to make sense of someone whose utterances and thoughts had the contents that these alternative generalizations anticipate. The construction of a T-theory based on those generalizations would be in violation of the Principle of Charity.

Starting with recipe a), once more I quote Davidson declaring that the principle “prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances” (Davidson, 1991, 211). *Rain near the speaker plus the fact that 2+2=4* - unlike simply rain near the speaker - is not a feature of the world the interpreter could likely see himself “responding to under similar circumstances”.

As for b), while again it is clear that this type of change would not affect the validity

of the generalizations, more needs to be shown to defend the idea that the alternatives could also allow for *charitable* content ascriptions. A grounded choice between the chemical descriptions and phenomenal descriptions - the examples L&L have to offer - seems perfectly attainable once the interpreter is allowed a full use of the Principle of Charity.

Background and context would be needed in each case to motivate any of these hypotheses, and, plausibly, it would also be enough to render only one of them acceptable: e.g. it is improbable that the member of some remote tribe showing no signs of scientific development should be talking about the atomic numbers of metals.

In c), L&L defend the possibility of exchanges between co-extensive nonsynonymous sentences. To illustrate the point, they affirm that the interpreter should be free to generalize about the proximal conditions prompting the speaker to hold the sentence true, instead of thinking in terms of distal ones. However, I defend that this alternative is not a genuine option for the interpreter. First, he does not know anything about such proximal events. No common human has any idea about what is going on with his or others' sensory receptors, no idea about the ongoing *patterns of irradiation* of his or others' *retinas*. Second, the interpreter is not expected to take such a possibility into account. The experience of radical interpretation, as Davidson thinks of it, is not meant to be so radical that the interpreter should even wonder if the speaker could not be talking about something completely outside the sphere of the interpreter's own experience, such as sensory stimulations. I believe this should result clearly from passages such as the following, where Davidson clarifies what resources he allows the interpreter to make use of:

> ... even my theoretically bereft interpreter has a lot more going for her than Fodor and Lepore seem to think. I have stressed that a radical interpreter already has a language, and a set of concepts that more or less match those of the interpretee. She has the concepts of truth, of intention, of belief, of desire, and of assertion (and many, many more). She knows a lot about the world and about how people behave in various circumstances. … interpretation can succeed only when interpreter and interpretee are much alike in important respects (alike in what they can perceive, alike with respect to built-in and learnable patterns of discrimination, alike even in size and degree of mobility). (Davidson, 1994a, 125)

The interpreter is expected to assume a lot. He is not to doubt or question whether he and the speaker share a basic worldview, whether his concepts *"more or less match those of the interpretee"*, and whether they are "*alike in what they can perceive"*.

Recipes d) and e) call for the same kind of response as the other recipes before. It

should be clear for the interpreter what must be included among the silent, background conditions, and what must take the prominent role. It is not the fact that the speaker is not having a stroke that is likely to be in her mind when she utters *s* while she points to the sky in disappointment as it begins to pour with rain.

This, I believe, is how an adequately strong understanding and employment of the Principle of Charity permits the interpreter to avoid arriving at an intractable number of alternative generalizations and, consequently, at underdetermination. It has been made clearer, I hope, how much harder than it is usually assumed to be it is to come up with a plausible case or example of underdetermination. I know of no credible example, and I doubt any can be found in this setting. This is my response to L&L's second challenge.

Without a solid case for underdetermination, the identified threats to the general conception of meaning and communication defended here are greatly reduced. Underdetermination of interpretation was noted above as a very significant symptom of disagreement between first- and third-person perspectives regarding the application of interpretational concepts, and as an obstacle to a Davidsonian middle path between Cartesianism and Radical Interpretivism. To severely narrow the plausibility of underdetermination is to severely narrow any evident danger of such disagreement, and to clear the way for the notion of theoreticity proposed above. In result, I maintain, it is still plausible and defensible to claim that there is no more to meaning than what can be interpreted by a vulgar interpreter in good enough conditions, essentially those delineated by Davidson for his radical interpreter, and that interpretability via rationalization determines meanings.

3.4. DIFFERENT CONCEPTIONS OF RADICAL INTERPRETATION

My disagreement with L&L regarding the tenability of the adoption of alternative generalizations, and consequent underdetermination, can easily be traced to a different understanding of the setting and goals of the enterprise. Lepore and Ludwig take what I see as an excessively austere and ambitious view of the experiment.

This austerity emerges in particular in their inventory of the kind of constraints the radical interpreter should be allowed to explore. At a certain point, they note that "[i]n practice, no doubt, we [regular interpreters] make many empirical assumptions about

speakers” (L&L, 2005, 219), and they specify what kind of assumptions they have in mind:

> We assume we and our subjects are very much alike in our basic and recurring desires, what we find salient in our environments, what we are apt to notice, what we find interesting or insignificant, kind or hateful, dangerous or comforting, and so on, always with a caveat about explicable differences, whose explanation would be grounded in an account of how we would, or might, have reacted to their history and situation. (ibid.)

But they reject that “the radical interpreter [should be allowed to] help himself to these sorts of assumptions” (ibid.). They offer two closely related reasons for this rejection.

First, they are committed to a very strict understanding of the idea that “the radical interpreter is ultimately to start from purely behavioral evidence” (ibid.) that precludes the adoption of those assumptions. “[W]hat a speaker is apt to find interesting, what his desires are, what he finds dangerous, attractive, and so on,” are not behaviorally given. So, they defend, the interpreter must either drop the assumptions, or else show how to derive them from more primitive data - i.e. show them not to be assumptions anyway, but acquired knowledge, thus rendering them an inessential, intermediary stage.

Second, admitting those assumptions, L&L believe, would compromise the generality of the exercise. They defend that “the project of radical interpretation” should aim “to be a completely general account of interpretation, and, hence, of meaning, communication, and whatever is essentially connected with these” (ibid.), and they fear that the adoption of the assumptions would limit the validity of the results and conclusions extractable from the experience to only “a particular group of speakers or kind of speaker” (ibid.)[71].

[71] Bjørn Ramberg makes an apparently equivalent point in the following passage:

“If the principle of charity is to serve radical interpretation, we must be extremely cautious lest we use it as a cover for theoretical shortcuts. Since it is true that little communication would actually take place without all kinds of social, psychological and linguistic conventions and presuppositions firmly in place, a methodological principle for actual interpretation would inevitably smuggle in assumptions that would obscure the semantic content of the theory of linguistic understanding. Since we are after a theoretical concept of linguistic meaning, we must try to prune our methodological principles of such labour-saving heuristic aids.” (Ramberg, 1989, 73)

However, he then goes on to assert that the possibility of radical interpretation is dependent on a certain idea of “pre-theoretical” and “absolute”, or “trans-linguistic”, understanding of truth. In doing it, I believe, he is taking in enough ingredients to ensure the sort of substantive common ground among speaker and interpreter that I affirm to be necessary for interpretation:

“… interpretation from one language into another works only because the interpreter possesses a pre-theoretical understanding of truth. That is, it works because the interpreter knows how to apply the truth-predicate of her own language. (...)

For example, as a speaker of English she has this understanding if she knows that "'snow is white" is true if and only if ‘snow is white’ is appropriate in a way that ‘snow is white’ is true if and only if ‘grass is

The two questions are naturally related. The reason behind a choice for behavioral evidence is, of course, its public and objective availability, simply evident and unassailable by the type of doubts that can be raised against the psychological assumptions here in question. But to strive for a deeper and more solid justification is also to strive for more generality and impartiality, and so the two motivations converge.

Altogether, this is the excess of ambition to which I referred. As I see it, to assume that the radical interpretation experiment is about an ideal position of neutrality and about ideal epistemological foundations, to assume that it is about how a thoroughly dilapidated and unbiased interpreter could still secure interpretative success, is to misrepresent and to exaggerate its ambition.

I do not think we should see the radical interpreter as purged of all *prejudice* and preconceptions. I defend, in particular, the thesis that the radical interpreter should, to some extent, be allowed to employ what Lepore and Ludwig exemplify in the previous passage as *empirical assumptions*. I also defend that Davidson thought it so himself. I will start with the latter, exegetical point, and then turn to the defense of the position.

Although sometimes he may have suggested otherwise, Davidson's point was never

---

green' is not - even though both sentences are true. What she knows is that for a speaker of English the belief expressed by the sentence 'snow is white' is caused by the fact that snow is white, not by the fact that grass is green.(...)

(…) It is this very intuition of truths-for-languages as somehow the same that drives interpretation. Davidson's strategy works by holding truth constant between the languages; T-sentences serve to calibrate TL and L precisely because while the Tarskian characterization of truth is always the characterization of the truth-predicate for a given language, thus bringing out its semantic structure, truth is not relativized to a language. The concept of truth that underlies a theory of interpretation is a concept of absolute truth. That is to say, in a true T-sentence, s and p are appropriate to the occasions of empirical observation in the same manner. It is by assuming this sameness of truth, which is the intuitive foundation of Davidson's model of interpretation, that the interpreter is able to understand L. It is by virtue of this trans-linguistic notion that she is able to formulate an empirical theory that in specifying how the truth-conditions of sentences of L are determined by their parts - that is, in characterizing the truth-predicate of the language - actually interprets the language." (Ramberg, 1989, 75-7)

The rejection of the idea that Charity commits the interpreter with any form of parochialism results more emphatically in Ramberg's articulation of the matter. However, at this stage, my concern is the opposite - that is, the excesses of a boundless and blank *universalism* (for lack of a better term). To be sure, I absolutely agree that Charity is not meant to sanction any form of indulgence in prejudices and biases specific to the interpreter's civilization, culture, community or tribe. The interpreter should be free and flexible to reach beyond such circunstancial limits when the opportunity requires it. At the same time, however, Charity recognizes that there is a common human and rational form, and that the interpreter has no option but to start within his own story and with his own standards. I believe that this is what Ramberg incorporates in his picture, when he stresses the requirement that speaker and interpreter should be able to converge on the same truths; that is, when he requires that they have the semantic resources to form sentences that are "appropriate to the occasions of empirical observation in the same manner".

to fix the interpreter into a position of radical impartiality, a cosmic exile. Davidson said many times, and in many ways, that interpreter and speaker must, to some degree, “share their way of life” (Davidson, 1995b, 51) and that “understanding depends on finding common ground” (ibid).

From an abundant number of passages where Davidson incites the interpreter to use his own standards of rationality, broadly taken, and, to some extent, his own vision of the world in the interpretation of his fellow humans[72], there is one in particular I want to return to here in a moment because it so flagrantly contradicts what Lepore and Ludwig defend for the radical interpreter.

As a possible way of limiting the range of alternative generalizations and subsequent

---

[72] Here are some of these passages to add to those already quoted above:

“To understand the speech of another, I must, be able to think of the same things she does; I must share her world.” (1982, 105)

“… the key to the solution for simultaneously identifying the meanings, beliefs, and values of an agent is a policy of rational accommodation, or a principle that Quine and I, following Neil Wilson, have called in the past the principle of charity. This policy calls on us to fit our own propositions (or our own sentences) to the other person’s words and attitudes in such a way as to render their speech and other behavior intelligible. This necessarily requires us to see others as much like ourselves in point of overall coherence and correctness - that we see them as more or less rational creatures mentally inhabiting a world much like our own. (...)

In the case of belief, what insures that our general picture of the world is one we share with other thinking creatures, and one that is, in its main commonsense features, correct, is that sentences, and the thoughts they may be used to express, are causally tied to what they are about. For in the plainest cases we can do no better than to interpret a sentence that a person is selectively caused to hold true by the presence of rain as meaning that it is raining.

… since the objects of your beliefs and values are what cause them, the only way for me to determine what those objects are is to identify objects common to us both, and take what you are caused to think and want as basically similar to what I am caused to think and want by the same objects.” (1984a, 35-36)

“... a radical interpreter already has a language, and a set of concepts that more or less match those of the interpretee. She has the concepts of truth, of intention, of belief, of desire, and of assertion (and many, many more). She knows a lot about the world and about how people behave in various circumstances. … interpretation can succeed only when interpreter and interpretee are much alike in important respects (alike in what they can perceive, alike with respect to built-in and learnable patterns of discrimination, alike even in size and degree of mobility).” (Davidson, 1994a, 125)

“I have described in its most transparent form the art of applying the formal theory to an actual individual, with both interpreter and speaker outfitted with a mature set of concepts and the linguistic aptitudes for expressing them. All that is lacking at the start is a shared language, and prior knowledge of each other’s attitudes. » (1995a, 129)

“I ... know an infinity of things that I can express and which I know someone else might believe or doubt or wonder about; the list is in a sense as large as the list of things expressible in my language, in concepts I command. These are the propositional contents to which I advert when I attribute attitudes to myself or to others. This rich reservoir of conceptual resources is what I must use in interpreting the utterances or actions of those around me.” (1998, 89)

interpretations - discussed above - Lepore and Ludwig consider, but ultimately reject, an appeal to something they name “the principle of saliency” (L&L, 2005, 234).

The Principle of Saliency: “ceteris paribus, take the speaker to find salient in his environment what you find salient, and to be thinking about or noticing what you would in his place” (ibid.)

They take the principle as a paradigmatic *empirical assumption* and reject it for the reasons presented above. But compare it with the passage from 1991, “Three Varieties of Knowledge” - already repeated a few of times - where Davidson explains that the Principles of Charity “prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances” (1991, 211). In refusing the interpreter the “*principle of saliency”* L&L are refusing him something virtually indiscernible from the Principle of Correspondence as Davidson himself characterizes it.

Now, for the second question. What reasons can be invoked to defend this position - the less austere stance on radical interpretation that allows the interpreter to liberally project his own standards in the understanding of his peers? My argumentation will mostly be negative. Its core will consist of the briefest sketch of an attempted rebuttal of L&L’s reasons presented above against this less austere stance.

Regarding the first reason - that to assume general homology between the interpreter and the speaker’s mind is to fail to meet an adequate standard of justification - my claim is that the project of basing such an investigation exclusively and integrally on behavioral evidence, without the benefit of some sort of assumption, is a misguided ideal. This takes us back to the arguments developed above, in §2.2. and §2.2.1. Behavior by itself is perfectly silent about the agent’s states and meanings; it needs an adequate framework to become minimally revealing. To demand of the interpreter a full grounding of his meaning ascriptions based solely on the neutral observation of the speaker’s behavior is to demand too much. Exorbitance, however, is not the requirement’s only fault. There also seems to be some inconsistency to it. How can one start by assuming that the other is a thinker and a speaker, that she is engaging in linguistic practices similar to our own, that she tells what she believes to be the truth, that she uses her language consistently, that she is collaborating with us, that she has regular logical and decision theoretical skills, and then wonder whether she

experiences and organizes the world in the same way that we do or whether she does it in some radically and irreparably different fashion? This mixture of realistic acquiescence in stage setting and strict requirements afterwards makes for a very artificial interpretation problem. This, in turn, speaks against the austere conception of radical interpretation that takes the experiment to be staging it.

I also dispute their second reason - that the adoption of the assumptions in question compromises the generality of the experience. To start with, take Lepore and Ludwig's rejection of a *Principle of Saliency*. In support of its dismissal they affirm that "it is certainly not conceptually necessary that a thinking being find salient or notice or think about the same things in its environment that we do" (L&L, 2005, 235). Davidson, however, provides us with enough elements to question their refusal.

First, following Davidson, we can try to deflate the notion of *conceptual necessity* to a more manageable *size*. About the special status[73] he claims for some of his assertions, Davidson explains that such convictions

> …can seem either empirical or a priori; … Empirical if you think it just happens to be true of us that this is how we come to be able to speak and think about the world; a priori if you think, as I tend to, that this is part of what we mean when we talk of thinking and speaking. After all, the notions of speaking and thinking are ours. (2001a, 294)

Second, in his defense of Charity, Davidson frequently invokes the fact that the interpreter would not be able to ascribe thoughts and meanings to a creature diverging from him too radically in her perceptual experience. This, in turn, should challenge him in his conviction that she talks and thinks. Being somewhat more realistic and restrained in the assessment of what we are actually ready to do with our notions, of what kind of occasions can genuinely call for the application of our concepts - the kind of intermediate stance between the *a priori,* the speculative and the empirical that Davidson hints at - should bring us to accept that a requirement of perceptual agreement between all fellow thinkers and speakers is not so farfetched after all.

Saliency is just an example. The same reasoning is extendable to all agreement that is needed for interpretation. A lot is needed, even if, as noted above, one should not hope for a

---

[73] He seems to use all of these - *a priori*, conceptually necessary, constitutive, and more - more or less interchangeably - see §.2.1.1. above.

precise delineation of what is clearly indispensable and what is only accessory. And the agreement might not be obvious or ready-made. It might take some bending and turning of the customary views; it might require some flexibility and invention.

Allowing the interpreter enough familiarity with the speakers' minds does not have to impede reflection on radical interpretation from reaching general "conclusions about meaning and language as such" (L&L, 2005, 219). All agents, thinkers and speakers ought to be much alike in diverse respects. Even when we delve into more substantial and concrete matters of our mental lives, we can still be talking about general conditions for agency, thought and speech[74].

3.5. CONCLUSION

In this Chapter I have presented and resisted two challenges presented by L&L against the radical interpretability of thoughts and meanings. My response to these challenges involved a particular understanding of Davidson's views on meaning and interpretation, as well as of his central thought experiment - radical interpretation.

In the first part, I rejected an Interpretivist understanding of Davidson's position, together with the idea that he is committed to the rejection of subjectivity, and defended an alternative picture that allows for the reconciliation of first-person knowledge and the publicness of thoughts and meanings. I highlighted the role that the idea of triangulation plays in making room for this ingenious solution between Interpretivism and Cartesianism.

In the second part, against the hypothesis of abundant and easy to prove *underdetermination* of interpretation, I stressed the often underestimated power and scope of

[74] In the end, it must be acknowledged, it is unlikely that anyone has ever disputed that Radical Interpretation, when understood in the terms I have been delineating, is possible. The much more liberal perspective I am offering, recognizing an immense weight and power to the Principle of Charity, and a very wide notion of rationality, *almost* trivializes the claim. In all fairness, even the undisputed assertion of our natural ability to interpret others is not so perfectly trivial. The crucial idea, that our interlocutors' meanings are not actually given away by the available evidence, but demand an interpretive process that is in part ungrounded and creative, preserves some of its perplexing power even after the banality of the accomplishment is acknowledged. There can be no independent confirmation allowing me, at any point, to suspend my assumptions and to ascend to a fuller objectivity in my ascriptions. As Davidson declares: "the standards of rationality and reality on which I depend in understanding others are my own, and there can be no appeal beyond them" (Davidson, 1998, 90). Even to reflect upon, test, discuss or even refine those assumptions, the interpreter must already rely on them, "there is no escape from the fact that we cannot check up on the objective credentials of the measure we are using" (Davidson, 1995a, 133-4); see also Davidson, 1998, 90.

the Principle of Charity. In the final section, I outlined a defense of the heavy appeal to Charity and to an ample conception of shared rationality involved in this second response.

4. Rationalization and Language Acquisition

We turn now to empirical accounts of the actual process of language learning or acquisition. Radical Interpretation attempts to model, to some extent, this acquisition process, but there can be no doubts about the difference between approaches. There is no question about the distinctiveness between a project of rational reconstruction, as we might describe it, exclusively concerned with the logical nexuses between elements and moments, and the experimental approach attempting to describe and understand the real cognitive processes involved. Even so, my aim in this chapter is to reveal affinities between the two. In particular, I want to show the role that rationalization and the Principle of Charity actually play in actual cases.

In Chapter 2, I concluded in favor of the validity of the principle relative to the effectiveness of successful interpretation or communication. If indeed we communicate - which I am taking for granted in this discussion - things must be as the Principle of Charity affirms them to be: we must be sufficiently coherent and correct[75] in our thoughts, meanings and actions. This, as affirmed then, should be enough to sanction appeals to the principle as a legitimate method of interpretation. That is, if no interpretation can reveal agents that are not massively rational anyway, the interpreter must be free to provisionally assume the rationality of the agent in conducting his inquiry, an inquiry that must, in the end, confirm that condition.

In this chapter I want to take these conclusions a bit further. I hope to show that moves and strategies that can aptly be described as instances of appeal to the Principle of Charity are effectively employed by children learning their first language. Instead of stopping at the conclusion that rationalization is a licit strategy for language learners and interpreters, I am trying to show that it is also actually used in the most common language learning situations, that is, I am trying to show that children employ rationalization in the inquiry they

[75] In the arguments above I aim only at *agreement*, not actual truth. Agreement is all that is needed in support of my thesis, which is why I can neglect the distinction here.

must carry out to become competent speakers.

4.1. RATIONALITY AND RATIONALIZATION

Rationalization of an agent, thinker or speaker, the ascription of meanings and mental states in accordance to the Principle of Charity, makes no sense outside a context of rational inquiry. The rationality of the process of language acquisition is a necessary condition for rationalization to play any role in it. Consequently, I am sustaining with this thesis that it is appropriate to describe language acquisition as a rational process.

It is maybe the case that every rational process is also susceptible to being described in other terms, and that rationality is realized by, or supervenient on, properties of other and more basic types. If this is so, the best question to ask may not be whether an appeal to rationality is required to account for the acquisition of language. A better question would be if such an appeal is suited to illuminate any particular aspect of that task, even if not the whole of it. Again, I believe the answer to be positive.

When what we are dealing with is the process of language acquisition, by calling it *rational* I mean that such a development is *in part* - it is important not to miss this qualification - aptly described as a process akin to that of belief formation, guided by, or in conformity to, principles of what is commonly referred to as *theoretical* or *epistemic rationality*[76]. To come to know or believe that one's interlocutor means *that so and so* by the use of some particular expression of hers, the learner must derive this piece of knowledge or belief in an adequate fashion from some adequate epistemic source.

In the course of the immense and complex task of learning a new language, the learner will find opportunities to employ diverse types of reasoning and to rely on diverse types of sources. For instance, not only will the learner, so to speak, half *induce* half *abduce* the referential intention behind some speaker's use of a certain expression e_1 from both *present* and *remembered observations*, but the learner will also consult *reason* within him and mobilize some *deductive reasoning* to infer from some premises that the speaker must be meaning a certain logical operator by her use of e_2.[77]

---

[76] See, for instance, Robert Audi, 2004.

[77] This is merely illustrative, i.e. in mentioning only the three traditionally noticed types of reasoning I do not

The rationality of the process does not, however, require language learners to be explicitly aware of reasonings and principles. For sure, there will have to be some conformity to them, and it is likely that interpreters will be at least partially conscious of their epistemic efforts and progresses, but I am not interested in details concerning how and to what extent. I will not get into such issues as types of knowledge - practical *vs.* propositional, tacit *vs.* explicit, knowledge that *vs.* knowledge how - or modes of conformity to rules.

Moreover, of course, it is not the case that language learners should be perfectly or optimally rational; there is room for more or fewer faults and mistakes, more or less talent, inspiration and luck.

I am being mostly vague and open in this characterization so far, but that is befitting as that is the type of claim I want to advance here in affirming that there is something worth calling *rationality* at work in the context of language acquisition. I aim at no more refinement or detail, in particular because we need the notion apt to be applied to the thoughts and actions of very immature learners also. Even so, from the beginning there is something distinctive to these processes - at least to some parts of them - something for which we posses no better word or description than *rational*. In a similar spirit of minimal commitment, here is a passage from Paul Bloom's book *How Children Learn the Meanings of Words* - a pivotal text in the recent study of language acquisition - distinguishing what counts as learning from what does not.

> The notion of learning picks out a subset of environmentally caused events, those in which the organism comes to store and represent information through a rational process ... of interaction with the environment. The caveat of ''rational'' is present to capture the intuition that not any interaction counts: if you get smacked in the head and miraculously come to know the rules of baseball, this wouldn't count as learning. But if you come to know baseball by observing other people play the game or by having someone explain the rules to you, then this does count as learning—even though, of course, this process would be impossible without the innate ability to learn. (Bloom, 2000, 15)

And he uses this passage precisely to introduce his own assertion that there is learning and rationality in language acquisition, more specifically, in word learning - this is also where I will zoom in on this chapter's question.

> This is a crude definition, but it captures the sense in which word learning counts as learning.

---

mean here to suggest any type of limit on the type of rational processes admitted.

> In fact, word learning is the clearest case of learning one can imagine. Nobody was born knowing the meaning of the English word *rabbit.* Everyone who knows the word has heard *rabbit* used in a context in which its meaning could be recoverable from the environment using a rational process; that is, everyone who knows the meaning of *rabbit* has learned it. (ibid.)

4.1.1. RATIONALITY IN WORD LEARNING

Were language acquisition only something that merely happened to humans, a kind of growth or maturation of the organism, much like visual development or the progress in motor skills, then this chapter, and this dissertation, would be pointless. Consider, for instance, the inappropriateness of inquiring over the rational aspects and nature of the acquisition of walking. Yet it is far from uncommon to find prominent figures in the study of language who are willing to explore and promote analogies between language acquisition and precisely such dumb and natural processes. See, for instance, how Steven Pinker puts it in his *The Language Instinct.*

> Language is not a cultural artifact that we learn the way we learn to tell time or how the federal government works. Instead, it is a distinct piece of the biological makeup of our brains. Language is a complex, specialized skill, which develops in the child spontaneously, without conscious effort or formal instruction, is deployed without awareness of its underlying logic, is qualitatively the same in every individual, and is distinct from more general abilities to process information or behave intelligently. For these reasons some cognitive scientists have described language as a psychological faculty, a mental organ, a neural system, and a computational module. But I prefer the admittedly quaint term "instinct." It conveys the idea that people know how to talk in more or less the sense that spiders know how to spin webs. (Pinker, 1994, 18)

Chomsky, of course, is a - or even *the* - principal figure in this line of thought, i.e. in defending the naturalization of language and the study of language, and in refusing to see its acquisition by humans as an exercise in theoretical rationality. He speaks – or at least spoke - of *knowledge of language,* but when he does, he uses *knowledge* in a technical and peculiar sense. In his 1986 book, *Knowledge of Language,* he explains that "to know the language L ... is to be in a certain state S_L of the language faculty, one of the components of the mind/brain". (Chomsky, 1986, 221) The acquisition of such *knowledge* is described, in terms the least susceptible to invoking the presence of any active intelligence, as *parameter setting* (and *periphery formation*). At one point, decisively relinquishing any familiarity with more traditional notions of knowledge or learning, Chomsky adds:

> The result of this process of parameter determination and periphery formation is a full and richly articulated system of knowledge. Much of what is known lacks relevant grounding in experience, justification, or good reasons and is not derived by any general reliable procedures. (Chomsky, 1986, 222)[78]

This should give us some pause for thought. What have we here? Is the topic this controversial after all? Are there prevalent views radically incompatible with seeing language acquisition as a rational process? I do not think so, not in the mitigated sense with which I am putting forward the claim that there are *aspects* of language acquisition that can properly be understood as rational processes - but not the whole of it, not most of it, but essential parts of the process nonetheless. What I want to depict as rational and involving rationalization is not the development of syntax or morphology, let alone phonology, nor even the acquisition of concepts or meanings. Instead I will concentrate on a different achievement, perhaps modest by comparison, but still indispensable for communication and understanding. To simplify, let us call it the learning of word meanings, that is, the learning of which meaning is linked with which primitive expression of a certain language or idiolect. Here, I would claim, rationality and rationalization are clearly called for. I hope this much can be accepted and accommodated by everyone.

Paul Bloom, for one, voices his conviction that there is consensus over the issue.

> There are two facts about word learning that everyone accepts. The first is that words really have to be *learned.* There is controversy over how much conceptual structure and linguistic knowledge is innate, but nobody thinks that this is the case for specific mappings between sounds (or signs) and meanings. (Bloom, 2002, 37)

Barry Smith gives us another example in his 2006 text, “What I Know when I know a language”, when he addresses precisely this discrepancy between types of knowledge involved in linguistic competence, and their respective acquisition processes, in the special section he reserves in it for *‘Knowledge of Word Meaning’* (Smith**,** 2006, 978-80). After spending the bulk of his long critical review of these subjects and polemics attacking typical philosophical views of language and the depiction of language acquisition processes as rational, theoretical or inductive, in favor of cognitive conceptions in the line of Chomsky,

[78] Despite the frequent convulsions in Chomsky’s views and research programs, he retains now more or less the same views concerning the aspects of language acquisition he is referring to here - that it should be seen as a process of biological maturation not as one of rational learning - as he did in 1986.

Smith is still careful enough to mention this important exception:

> Unlike our knowledge of grammar, there is no reason to think that our knowledge of word meaning is inaccessible or sub-personal. (...)
>
> Our knowledge of word meaning is conscious and first-personal. There is such an experience as the meaning of a word being all there at once, or of bringing the meaning of a word to mind as when one decides whether the use of a particular word is more apt than another. These experiences of meaning belong at the personal level. How do we acquire them and how can we use them to understand others? The quick answer, that can only be sketched here, is that we learn to have experiences with words in the context of learning words from others. The early conditions for word learning typically happen best under conditions of *joint attention* where the child and the parent are jointly attending to a commonly perceived object. The sharing of their experience of that object can be commemorated by introducing a sound label that saturates the experience. (...)
>
> The combination of these two systems — for word meaning and for syntax — brings about a dimension shift in the expressive power of the language user. Combining such knowledge is necessary for full language acquisition. But what we see is that the experience of meaning and the experience of hearing strings as structured respond to different parts of cognition and despite the experience of hearing what you say as there in the words uttered, the sources and objects of these two kinds of knowledge are quite different. (Smith, 2006, p. 978-80)

We can also confirm a general amenability to our claim - that we can and must appeal to rationality in accounting for some aspects of language acquisition processes - by those very same authors with whom the fear of controversy first started a couple of pages back.

Notice, for instance, that a hundred pages further on in the same book quoted above (Pinker, 1994, 149 ff.), Pinker turns his attention to lexical building. There, in sharp contrast to the passage quoted above, he is quite at ease talking about the '*learning'* of word meanings, comparing it to the '*memorization of facts'*, and elaborating on the role of '*induction'* in the process.

Chomsky, by recognizing the arbitrariness of association between '*concepts and sounds'* and the corresponding free variation between languages in this particular respect, also seems obliged to concede the special character of this extra piece of necessary knowledge that, unlike most of the rest, is left especially undetermined by human genetics.

> ... language variation appears to reside in the lexicon. One aspect is "Saussurean arbitrariness" the arbitrary links between concepts and sounds: the genetic program does not determine whether *tree*, the concept, is associated with the sounds "tree" or "Baum". The linkage of concept and sound can be

acquired on minimal evidence. (Chomsky, 1995, 121)

He is still rather elusive about the methods through which humans come to acquire it[79], but there is a patent difference between being acquired and supported on '*minimal evidence'* - as Chomsky in this passage describes knowledge of word meaning - and lacking *'relevant grounding, justification or good reasons'* - as he, in the passage above, described knowledge of language, i.e. knowledge of grammar.

Lastly, I want to point out that Davidson also acknowledges the pertinence of this distinction between the ingredients concurring for linguistic competence, and the nature of their respective acquisition processes.

> I have been talking all along as if learning a first language depended on natural induction. I am aware that Chomsky and those in his thrall resist the application of this notion on the ground that basic grammar is wired in, and that early language acquisition is too rapid to count as learning or to require anything like induction. Let it be so. But whether or not we call acquiring them learning, we are not born knowing the words of Bantu or Armenian. It may well be that one hearing is enough in many cases to give us all we need in order to 'go on' in the right way. This cannot change the point I am making, provided the one hearing is in the presence of an ostender who is ostending an appropriate object, event, property, or state-of-affairs.' (Davidson, 2001c, 15)[80]

In this passage he is talking about first language acquisition, not about Radical Interpretation, our prototype of a rational and rationalizing approach to interpretation and language acquisition. Despite the inevitable promiscuity between the two topics, we must carefully distinguish what Davidson has to say about each. Most importantly, Davidson sees the acquisition of the first language as a process inextricable from the emergence of thought itself, and hence altogether as a much more ponderous achievement. Yet Davidson himself asserts that they are not without some important commonalities. In the following passage, Davidson suggests that it is in part as if the first comprehended the latter, plus much more.

[79] See also Chomsky, 1992, p.61, Gleitman & Fisher, 2005, Gleitman, 2010, and Chomsky's reply to Gleitman in Chomsky, 2010, 394-7.

[80] See also this other passage from *Radical Interpretation Interpreted*: "I hold that the speakers of natural languages can be, and often are, correctly understood on the basis of non-linguistic facts not merely available, but readily available, to the likes of you and me. This view does not entail that we may not have a universal grammar wired in, or even that we don't have the grammars of five or fifty thousand languages wired in. But it does embrace the conviction that the references of unstructured singular terms and predicates are not wired in; if we understand a speaker, we know how her words are connected to the world, and this is something we cannot be born with. These connections are established both for the speaker and the interpreter in the context of social interaction." (Davidson, 1994a)

> Learning a first and learning a second language are, of course, very different enterprises. The former is a matter of entering the domain of thought for the first time, the latter is a matter of someone already at home in the realm of thought entering into the thought of someone else. Both, however, depend on similar mechanisms and similar cues**.** Furthermore, the contrast is weakened by the realization that in the case of the child initiate, the two forms of learning mesh, for in absorbing the idea of an objective world, the child is simultaneously learning to communicate with others, which requires insight into the thoughts and intentions of those others. (Davidson, 1998, 88)[81]

Despite the serious differences between the learning of the first language and the learning of a subsequent one, the fact - I defend - is that Davidson's prescribed method for radical interpretation still runs very close, at moments, to the actual processes of first language learning carried out by young children. This affinity, as I have been anticipating in this section, and will continue to defend in the rest of this chapter, is most noticeable with respect to word learning.

Contrast it, for instance, with the development of grammar. For all the structure that the radical interpreter might be able to recover by cleverly analyzing utterances, he could plausibly still be very far from supporting anything close to a full grammar. According to Chomsky's argument, owing to the poverty of stimulus it would be impossible to learn a grammar that way[82]. Even more to the point, experimental studies do not show any evident parallel - to say the least - between the radical interpreter's exercise of analysis and truth-theory building, and the processes of grammar learning, or maturation, in children.

The opposite is the case with word learning. The same elements pointed to as fundamental to radical interpretation - and also, derivatively, to all interpretation[83] - are also present and central to children's learning of word meanings. Just like the radical interpreter, children learn their words by taking into account the systematic relations between what is perceptibly the case and occasion sentences held true by the speaker, and the need to ascribe meanings and thoughts to the speaker so as to render her mostly rational - in the broad sense described in the previous chapters, and having the learner's own nascent standards as the reference. This is the predominant view in the field, and I will dedicate the next sections to its exposition revealing in some detail how the empirical study of word learning provides

---

[81] See also Davidson, 2001a, 294.
[82] See Chomsky, 1965, pp. 47 ff., and Chomsky, 1986.
[83] See the last section of the first chapter, 1.2.5.

support to Davidson's foundational account of meaning and communication.

4.2. EMPIRICAL ACCOUNTS OF WORD LEARNING

How do children learn the meaning of words? Unsurprisingly, it is assumed that in trying to answer this question we must place our initial focus on those situations where the infant attends to speakers' uses of words that are perceptibly and correctly related to what is going on around them. Just as Davidson naturally assumed in his radical interpretation, the first steps of the learning process must take place on those occasions where language and world come closest to each other, like simple references to observable objects and events. In fact, probably most texts on children's lexicon acquisition do not fail to mention the *gavagai episode* - an emblem of Quine's radical translation experience from which Davidson's Radical Interpretation followed - as an exemplary point of entrance to the process: a rabbit runs by, the speaker utters 'gavagai', the learner must figure out which apparent aspect of the situation prompted the utterance[84].

These are the propitious situations. But how are children to explore and learn words from them? In the recent literature on the topic[85], we find appeals to three main types of mechanisms[86]: associative learning, lexical principles or constraints, and social, pragmatic or mindreading skills. In the next sections I will briefly describe what mechanisms these are and consider their role in the process of word learning.

Naturally, none of these mechanisms are expected to work autonomously from a background of other maturing cognitive structures and abilities. They are put forward as, say, the closest contributing factor in the achievement of word learning, while other necessary ingredients, like the acquisition of phonology, and the development of the conceptual repertoire, are simply assumed and left unexplained in this section's discussion.

The respective weight and prevalence of each mechanism in the process of lexicon

---

[84] I find it a bit of a wonder that Quine gets mentioned so much in this literature and Davidson not once - as far as I could verify - even though he picks up the same problem and endeavors to find a type of solution that is much more congenial to some of the proposals currently enjoying ample endorsement.

[85] See, for instance, Golinkoff et al. (2000); chapter 3 of Ambridge & Lieven (2011); Bloom (2000); Diesendruck (2007).

[86] The most important omission in this list is syntactic bootstrapping. In this category fit all forms of reliance on grammatical cues that are used by the child in the process. I leave it out as it cannot aspire to be more than an auxiliary mechanism, and it cannot really be defended as an alternative to mindreading and rationalization.

acquisition is a matter of hot debate in the field. Different authors attribute different roles and importance to each. Some try to make do with only one or two of them, dropping the remaining. Proposals have also been put forward to explain effects first attributed to some mechanism in terms of the workings of some other. My aim here is not to contribute in any way to the progress of the debate or the solution to these polemics. I want simply to report the wide agreement behind the idea that mindreading plays an essential role - stronger or weaker, but indispensable - in the process, and to reveal some of the reasons and studies behind it.

Social cognition, *pragmatic understanding*, *mindreading* or *theory of mind* - there may be occasional subtle differences between the notions these labels are used for in this context by the authors I will be dealing with in this section. However, I will mostly ignore them or, even better, assume that all these capacities are exercised together by the child. Accordingly, I will typically use any of these labels broadly as encompassing all forms and manifestations of the infant's ability to recognize and understand the other - the speaker, the parent, the teacher - as a rational and social agent, endowed with a mind much like her own. My word for it is, of course, *'rationalization'*[87]. My point here is to show that the prevalent

[87]I anticipate challenges to the identification of all these notions. In particular, many voices have expressed several types of reluctance towards the idea that rationality and assumptions of rationality have an essential role to play in humans' investigations about the minds of others. Here I want to briefly address these points.

First, there is the question of whether or not we are dealing with rational agents after all. I responded to this in Chapter 2, more specifically, in §2.1.4.

Second, there is the question of whether or not the presumption of rationality is a good method for mindreading, or the one effectively employed by actual thinkers. Some theorists have claimed that it is not. This has been the object of intense discussion for the last decades. Some theorists have understood appeals to the Principle of Charity in this context as an endorsement of a very specific and narrow methodology. They take the rationalizing approach as a competitor and alternative to other mindreading methods. Take, for instance, this passage from Alvin Goldman's *Simulating Minds*: "How is mindreading accomplished? In broad strokes, there are three competing answers: by theorizing, by rationalizing, or by simulating. The first approach (theory theory) says that ordinary people construct, or are endowed with, a naïve psychological theory that guides their assignment of mental states. The second approach (rationality theory) says that the ordinary person is a rationalizer. She assumes that her friends are rational and seeks to map their thoughts and choices by means of this rationality postulate. The third approach (simulation theory) says that ordinary people fix their targets' mental states by trying to replicate or emulate them. It says that mindreading includes a crucial role for putting oneself in others' shoes. It may even be part of the brain's design to generate mental states that match, or resonate with, states of people one is observing. Thus, mindreading is an extended form of empathy (where this term's emotive and caring connotation is bracketed)" (Goldman, 2006, 4). On this basis, Goldman proceeds then to point out the insufficiencies of the second method, ending up by concluding that rationalization can, at most, be one parcel of the story and, thus, not required to all mindreading.

In response to this and related challenges, I want to note that we do not need to - and, in fact, should not - understand rationalization and the appeal to the Principle of Charity as portrayed in the passage, the adhesion to a particular methodology. Instead we should insist that what is being proposed is just a commitment to a resulting overall rationality, one that allows different methods to participate in the building of that result. Davidson himself is quite ecumenical in his elaborations and descriptions of what constitutes an appeal to Charity. For instance, there are passages running close to what Goldman labels the *Theory Theory* approach, in particular when Davidson stresses the theoretical nature of interpretational concepts. Other passages bring him

perspective in developmental research acknowledges the central role of rationalization in the process of word learning. Naturally, at the beginning we will find only the most incipient forms of rationalization, as will become patent when we get to the studies mentioned. These are, nonetheless, the unequivocal beginnings of humans' ability to understand each other as fellow-minded and rational beings.

4.2.1. ASSOCIATIVE PROCESSES IN WORD LEARNING

Let us start with the simplest and plainest available type of account - association. Associative learning mechanisms first reached prominence in philosophical and psychological explanations in the work of the British empiricists - Locke, Hume and Mill - and then in Pavlov's proposals and experiments, and in the behavioristic paradigm in psychology that followed, culminating in the work of Skinner in the 1950s.

One can characterize these mechanisms in a very general way as an ability to recognize covariation among elements of the agent's experience. Applied to the learning of words, an associative account, in its basics, would run close to this: the child's listening to the word "rabbit" when intensely aware of the presence of rabbits - rabbits being the salient object in the scene - forms in her the link between word and meaning, or referent, that constitutes her competence with the word.

Undeniably, the detection of this sort of covariation plays a crucial role in children's word learning. One could not hope to start teaching a word such as 'dog' to a prelinguistic child with no dog in sight, or otherwise salient. The question is rather how complete an explanation an exclusive appeal to associative mechanisms can provide. The prevalent position in the field is that more has to be added to the account to make sense of children's accomplishments. On the side of those who defend that an exclusively associationist approach is unviable, see, for example, Bloom (2000), pp. 56-60, Ambridge & Lieven (2011) p. 101, Diesendruck (2007) pp. 270-2, Akhtar & Tomasello (2000), Tomasello (2003) pp. 82-4, Hirsh-Pasek et al. (2004). It is harder to find unambiguous representatives of the opposite side of the dispute, i.e. defending an exclusive associationist approach, but Smith (2000) and

---

close to *Simulation Theory*, as when he says that the Principle of Charity "prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances" (Davidson, 1991, 211). In any case, the end result should be rational enough - this is the unifying trace.

(2000b) at least come very close.

Here are some of the problems and shortcomings hindering pure associationism. Even regarding words referring to concrete, perceptible objects and properties, such as mom, dogs, red, hot, chairs and cookies, it is common for the infant to hear the word while she is not attending to the referent in question. There are studies supporting this claim but I would expect a moment's thought to be enough to convince one of its truth. For instance, quite often the infant will be looking at the face of the person talking to her. Other times she will be much more curious about the toy she is holding, or the dog barking in her vicinity, but this will not stop the parent from uttering words that do not match her interest. Adults do not always wait for the child to be focused on the right referent before they speak, not even the most diligent modern parent. Finally, apart from that, children hear words all the time that are not directed at them, while they are looking and paying attention to all different sorts of unrelated objects and events, and children do pick up words in this indirect fashion nonetheless. Sometimes they have no other option. In some cultures, adults hardly speak to children before they acquire some rudiments of language, but these children do not fail to learn the language (Lieven 1994).

Help from repetition is limited. It could be thought that the *occasional* mismatch - e.g. "ball" is uttered while the infant's attention is with the dog - could be diluted in a large enough data collection, and that in the long run the infant would naturally converge towards the right association. However, such an appeal to statistical learning based on accumulated observations does not look like a promising solution to the problem. Not, at least, just by itself. As Bloom (2000, 59) notes, this solution would be inconsistent - or at least, very hard to fit - with the way children deal with new words. Children are frequently fast, confident and accurate with the learning of a new word, instead of slow, cautious and error-prone as such an account would seem to imply. In fact, children can even learn the word without any instance of contemporaneity between word and perceptual attendance - as shown in some of the experimental studies that we will consider in a moment.

This is not to affirm, of course, that children do not normally take cues from cross-situational comparisons[88]. It is simply that this does not happen in the undirected, unconstrained kind of way one should expect if only associative mechanisms and statistical

[88] As emphasized, for instance, by Yu & Smith, 2007.

analysis were in force.

Even if - or when - the child happens to be looking at the right target, there is the much more complicated question of its *delimitation*. What portion of reality, or quality of that portion, is the word related to? Let us say that the infant successfully detects that what the father is saying has something to do with the cat in front of them. But now, is he talking about the particular animal, maybe calling it by its name? Or about some part of its body - maybe the funny tail that still moves and looks alert while the cat lies peaceful and still? Maybe the father is commenting on some quality or property, perhaps the cat's laziness. Or maybe he is just noting the event of its presence in the room. Maybe, instead, it has something to do with what the cat is doing, its rolling itself to sleep, or something it did a couple of seconds ago, like having finished its plate of food. Possibilities are endless[89], and associative accounts have trouble explaining how any particular aspect is determined to be the one to be associated with the word (Tomasello, 2003, 43-4). Saliency becomes an impractical notion here - either unfit to settle the issue, or too vague and permissive when so many and varied options are in confront.

Things get even worse of course with words referring to gradually less concrete and perceptible things. As referents grow less visible and touchable, association becomes less viable as a candidate explanation. This happens sooner rather than later. Even if we stick to children's very early vocabulary, we find a solid representation of such words. Bloom refers to a study into the lexicon of a number of 20-month-old infants that revealed that "only about half of children's nominals referred to basic-level object kinds; the rest referred to members of other conceptual categories, such as locations (beach, kitchen), actions (kiss, nap), social roles (doctor, brother), natural phenomena (sky, rain), and temporal entities (morning, day)" (Bloom, 2000, 59).

Finally, there is also the frequent accusation of there being no account on offer about how associative mechanisms are supposed to form the child with an adequate understanding of the general nature of language, and of communicative interchanges - an understanding that is necessary if the child is to realize that the sound in question is being used as a symbolic

[89] And here I am just listing very prosaic alternatives. What we have here, patently, is just a new instance of the same general problem that has surfaced several times above of how to bridge the gap between observation and content ascription. In previous chapters, we have stretched the examples - such as *tabairs* and *quus* - beyond all likeliness.

device, a word, with some arbitrary and conventionalized semantic function. The recognition of the utterance event as a form of participation in a general and very peculiar sort of practice - language - seems like a non-negotiable condition for something to count as a learning or understanding of the words in question. As Tomasello puts it, “the child must determine, first, the adult’s overall communicative intention and, then, the particular way or ways that the new word is contributing to that communicative intention” (Tomasello 2003, 84). And he adds that “this complex set of cognitive and social-cognitive processes is not accurately described by the simple term ‘association’” (ibid., 84).

Proponents of associationism sometimes complain that criticism often misses the point of the approach. They accuse critics of taking associative mechanisms as horizontal alternatives to other types of mechanism such as lexical constraints and mindreading when what they propose is, in fact, a different-level explanation. They claim that they intend to explain the same phenomena but at a more elementary and general stage. As Linda Smith puts it:

> These associative-learning accounts would be implementation versions of their parent explanations, versions that mechanistically specify the undefined terms of “links” and “maps” and “predicts” and “expects” that fill the parent accounts. Specifying such implementation versions of the other accounts in this volume is a useful goal, since each account captures real and important truths about children; about language; and about how children become, in such a very short time, truly prodigious learners of words. But, at present, each is currently underspecified, couched in undefined folk-psychological terms. Because of this, none is a direct competitor of the associative-learning account; however, each could, instead, be realized in terms of associative mechanisms. Put another way, the associative- learning account resides at a different level of explanation than do the other accounts in this volume. It seeks to go behind folk-psychological terms to specify the processes and mechanisms out of which “beliefs” are made. (Smith, 2000b, 173-4)

This suggestion, however programmatic for the time being, is still perfectly compatible with my views and plan here. In trying to show there is a general agreement in psychology and developmental sciences about the need to involve children’s understanding of others as social, rational and minded beings in an understanding of how they manage to learn the meaning of words, I hold no position regarding how that understanding is constituted. I make Bloom’s words mine when he declares that defending the essential place of mindreading – which he calls ‘theory of mind’ - “leaves open the possibility that the mechanisms underlying word learning, while themselves not associationist, are somehow the

product of associationist learning mechanisms. In particular, if a connectionist theory can account for the origin and nature of the relevant theory of mind capacities, then connectionism is consistent with the facts of early word learning. If it can't, it isn't." (Bloom, 2000, 60)

4.2.2. LEXICAL CONSTRAINTS OR PRINCIPLES IN WORD LEARNING

An appeal to lexical principles or constraints goes with an understanding of word learning as an inductive problem. The infant is represented as having to figure out the correct word referent or meaning from a logically infinite range of hypotheses, starting only with the minimal assumption that the word must refer to or describe some event, object or aspect of the present situation. The principles or constraints now in question are conjectured as forms of limiting the variety of relevant hypotheses to a more manageable number.

They are proposed in different forms: more or less strict, innate or acquired, acting in different types of combinations with other factors[90]. One crucial thing all proposals in this category must share is the specific and limited scope of the mechanisms: they are meant to govern only lexical acquisition.

A couple of examples are in order. The whole object bias is presented as responsible for a tendency in young children to assume that a novel word is being used to refer to a whole object, in contrast, for instance, to a part or property of an object, or to some activity in which it is engaged, e.g. the cat and not its tail, the white of its fur, or its sleeping. The mutual exclusivity bias is supposed to account for the tendency to assume that there is only one word per meaning. If the child already has a word for cat she will likely assume that the new word means or refers to something else. As we see here, several constraints or principles are allowed to work together and to interfere with one another.

There are several problems with this type of solution. First, some critics claim that it is doubtful that a limited and plausible number of such principles or constraints could relevantly narrow the number of hypotheses open to the child[91]. Second, children must often go against these constraints if they are to learn the words properly, which they do. Some

[90] Compare, for instance, Markman, 1989 and 1992, with Hollich et al. (2000).
[91] Ambridge & Lieven, 2011, 67.

critics[92] go even further to claim that such constraints or principles might even hinder more than help the learning of words.

However, the most interesting objection - from our point of view, at least - is the one that challenges the adequacy of postulating mechanisms at such a local level, i.e. specifically lexical. While the identified tendencies are indeed detectable, they might be better explained in different terms by means of more general mechanisms. In particular, proposals have been put forward to explain the same effects by invoking instead children's social and pragmatic understanding of the situation, and their familiarity with the workings of other people's minds.

For instance, instead of a primitive tendency in children to assume that new words are used to refer to whole objects, it has been proposed that children's apparent object bias is due to the fact that they already know enough about other people's minds to assume that when adults use words in *certain contexts,* they will likely be referring to objects, but also that in *certain other contexts*, they will likely be talking about something else – see Tomasello & Akhtar (1995), Tomasello (2003, 85-86) and Bloom (2000, 98). As for the other example, the thesis has been defended that children's tendency to expect a new meaning or referent for a new word is, again, rather the product of children's familiarity with the way their interlocutors act and think - see Bloom (2000, 65-70) and Ambridge & Lieven (2011, 77-9). We will return to this last example to elaborate on the reasons and studies that support the proposed revision, but only after we have become better acquainted with the last type of mechanism.

4.2.3. MINDREADING AND RATIONALIZATION IN WORD LEARNING

In the last few decades, cognitive scientists have come to acknowledge the presence of an ever growing number of varied and rich cognitive skills in the prelinguistic infant.

Tomasello (2003, 3) and Bloom (2000, 62-63) list a few. Starting at 9 to 12 months old, infants become sensitive to, and able to follow and deliberately influence, other people's attention. They can easily follow the mother's gaze and find out what she is looking at, thus engaging in situations of shared attention to common targets. They operate the other way

[92] For instance, Tomasello, 2003, 86-7.

around as well. At around the same age, they start exploring what rudimentary instruments they have at their disposal - gestures, including pointing, and vocalizations - to direct others' attention towards whatever they find of interest. They also reveal a teleological understanding of certain behaviors, and a nascent ability to learn by imitation the intentional actions of others.

A very interesting fact about these early abilities - gaze following, imitation - is that they seem to depend upon the child's recognition of her partner as belonging to the relevant category of animate and purposeful entity, something the child seems to "diagnose via a combination of surface characteristics such as eyes and fur and the entity's propensity to engage in contingent behavior" (Baldwin & Meyer, 2007, 91). Baldwin and Meyer refer to a couple of studies supporting these claims and explain that these findings are important because they support the idea "that gaze following and attention to gestures as early as 12 months represents a genuine form of social responsiveness, rather than a non-social instance of reflexive orienting" (ibid.).[93]

New and more complex abilities follow suit. For instance, in a 2006 study (Liszkowski et al. 2006) one-year-old children point to inform the adult of the location of an item that the adult had dropped unintentionally and seems unable to localize. This reveals the child's ability to understand the interests and goals of the experimenter, as well as his lack of relevant information, and her willingness to intervene and help. This is even more impressive as the object dropped in the experiments would be something dull and uninteresting, such as a marker or a hole punch, while other objects, more attractive toys, were also lying around.

Lastly, recent research[94] shows results clearly supporting the claim that children as young as 25 months old are already able to ascribe false beliefs to agents and to predict their actions in accordance. Previously, the prevalent view was that children younger than 4 years old did not have a secure enough grasp of the relevant mentalistic notions as they consistently failed the tests available. Later experiments with more sophisticated designs - in particular, involving eye-tracking technologies - revealed that this starts much earlier.

What we have here is a growing list of independently confirmed early abilities of mindreading and rationalization. Many authors propose that these abilities are also deployed

---

[93] On the same note, see also Bloom, 2000, 62.
[94] See Southgate, 2013, for a review.

in word learning. Word learning, the story goes, depends on the child's ability to recover the speaker's communicative intention. This is easiest in a context of joint attention between child and adult to a new, conspicuous object that the adult points to and names in a clear and redundant fashion. However, what is crucial is the inferring of the speaker's intentions, and there are other ways for children to achieve this. A convincing collection of studies, produced throughout the last two and a half decades, illustrates the diversity of clues children track and the complexity of means and reasoning they explore in this type of inquiry. I will briefly describe some of the most important.

In the early nineties, Dare Baldwin produced the first, widely cited, series of studies supporting the hypothesis that mindreading skills play a crucial role in children's word learning.

In Baldwin (1991), children between 16 and 19 months of age were trained in two distinct conditions and subsequently tested for results. In both cases children were presented with two novel, attractive objects - two toys. Next, the experimenter removed one of the toys and placed it inside a bucket hidden from the infant's view.

Then, in the first condition, *the discrepant labeling condition*, the experimenter handed the visible object to the child and held the bucket with the hidden toy in her hands. She waited for a moment when the child was looking at the visible toy before looking into the bucket and producing the new, made-up word: "Oh, it's a toma!". The new word was repeated four times, its utterance always accompanied by the peering into the bucket, and initiated at a moment when the child was paying attention to the visible object.

In the second condition, the *follow-in labeling condition*, the experimenter prepared the situation in the same way. She handed the visible toy to the child, and held the bucket in her hands, but in this case, instead of looking into the bucket, she gazed at the visible toy while she uttered the new word.

A testing phase followed, where children were shown both toys and asked questions such as "Where's the toma? Can you show me the toma?". The results are particularly clear for children in the older group, 18 to 19 months old. Around 70% of those in the follow-in labeling condition chose the visible toy, but only around 30% of those in the discrepant labeling condition did the same. This shows that even at this young age children are sensitive to clues regarding the adult's communicative intentions - in particular, the direction of the

adult's gaze during the utterance. Despite all the trouble involved - the referent was hidden, the wrong toy was visible and was the actual object of their attention at the moment when they heard the utterance - most children still managed to re-direct their attention to join that of the adult and to infer that the new word referred to the object she was looking at.

These results were confirmed in the following studies by Baldwin (1993a) and (1993b). The training conditions were refined in certain aspects and the results purified of certain ambiguities, but the general conclusions were the same: children use more than associative mechanisms in word learning and, from the beginning, exhibit the sort of attentiveness to their interlocutors that qualifies as mindreading.

In 1996, a new type of experiment was staged (Baldwin et al. 1996). While the studies in the first three papers were designed to show the role intention reading played in word learning, this one is more straightforwardly directed at confirming the insufficiency of simple associative mechanisms. All children were presented with a new object and, while they had their attention fixed on it, they heard the novel label: *"A dawnoo! That's a dawnoo!"*. However, a first group of children heard it from a speaker that engaged with them in joint attention to the object, while a second group heard it from a *decoupled* speaker - someone standing behind a screen and apparently participating in a different activity like having a phone conversation. Among those between 18 and 20 months old, only children from the first group managed to make a link between the new label and the new object, thus supporting the idea that mere contiguity between sound and attention to the object is not enough to learn the meaning of a word.

Tomasello, working together with a number of colleagues, produced a second series of very revealing studies. Some of these experiments involve a search game. In the first one[95], the experimenter declares her intention to find a certain object - *"Where's the gazzer? Let's find the gazzer."* - and goes looking for it in a group of buckets, each hiding a different new object. The experimenter extracts one object at a time and holds it for the child to see. She will then either show signs of discontent - when the object is not the right one - return the object to the bucket and continue the search in the next bucket, or show herself satisfied - when the *gazzer* is found - smile and stop the search. Children 18 and 24 months old learned the new word for the object the adult intended to name. It made no difference whether the

[95] In Tomasello & Barton, 1994, and Tomasello et al. 1996.

object was found in the first bucket, the closest to the utterance of the new word, or in the later ones. Thus, again, very young children revealed an adequate understanding of the situation and action performed by the experimenter, which clearly requires attention to and a good enough understanding of their interlocutors' states of mind.

The involvement of such mindreading skills in word learning situations is confirmed yet again in a related experiment[96]. This time the child is first instructed about the location of four new, different and unnamed objects. When the experimenter is sure that the child knows which object goes in which place, she announces her intention to *'find the toma'* and goes to the toy barn - the hiding place of one of the objects. Then, in the *referent* condition, children watched the experimenter extract the target object from the barn, and were then allowed to play with it. In the *absent referent* condition, the experimenter attempts but fails to open the barn and extract the target toy, explains with a frown of disappointment that *'it is locked'*, moves on to the other hiding places from where she extracts the other toys, smiles at them and hands them to the child.

In both conditions, children from 18 to 24 months of age managed equally well to grasp the intended link between the new word and the hidden toy that was never seen again, in the absent referent condition, between the time the word was first uttered and the testing stage. This confirms that word learning is not about forming associations between perceptible objects, properties or events. Instead it is about figuring out the speaker's communicative intention, a task for which they mobilize all their interpretive and mindreading skills, all their little, accumulated knowledge of the world, other people, their habits, practices, interests, social settings, routines, and institutions. The last two situations, in particular, are too rich and complex to imagine that the child might be able to make any sense out of it without a mentalistic understanding of the experimenter's behavior - by means, for instance, of some specific constraint.

As already noted above, besides the question of grasping the particular meaning or reference for each new particular expression, there is also the previous question of understanding the general nature of language and communicative practices. Here again it is doubtful the learner could even start distinguishing the relevant actions, and understanding what they were all about, without her possessing and exercising the type of social

[96] In Akhtar and Tomasello, 1996, and Tomasello et al., 1996.

understanding and mindreading skills we have been talking about. I am deliberately giving only sparse details so as not to commit myself here with any particular view concerning how to analyze a successful communicative event in terms of the underlying mental states of the participants[97]. However, the necessary presence of some ingredients seems beyond question.

There is, on the speaker's part, the necessary presence of some type of communicative intention, and some type of expectation regarding her interlocutor's ability to recognize that intention and of associating the right meaning with her words.

On the listener's part, the child's part, there should be a corresponding understanding of the speaker's intentions, expectations and meanings. First, she must understand the sound production as a deliberate action with a teleological structure and explanation. This requires notions - however rudimentary - such as *will*, *desired effects*, *means*, *beliefs*, *decisions* and *intentions*. But the child must go beyond that. She must also understand the specific nature of the act in question, namely, the peculiar way in which this type of action seeks to influence the interlocutor's mind or behavior, the coordination and cooperation involved. And still this is not all. There is also the conventional nature of these practices. From very early on, the child seems attuned to the arbitrary nature of the associations between the sounds and meanings she is presented with, and capable of grasping its point in securing valuable coordination among the members of an extended community.

Not all of this can be expected to be clearly there from the beginning in its mature and definite form, but much seems to be necessary to make sense of children's responses to the linguistic events they are exposed to. An impressively solid understanding of these practices and a grasp of the relevant notions is already patent in very young children. We could already see much of this in the studies presented above, but there is more, and of particular interest to this last cluster of topics, in the following two experiments described by Diesendruck and Markson (2001).

The main purpose of the first study was to support a decision between two competing accounts of the same effect: children's avoidance of lexical overlap. As explained above, children tend to assume one name per referent and, more generally, one word per meaning. For instance, when presented with two objects, a familiar one for which they already have a

---

[97] Say, for instance, in terms of the formation and recognition of a complex Gricean meaning intention (Grice, 1957).

name, and a novel one for which they do not have a name, and asked for the referent of a novel name, children tend to choose the novel object. Several studies referred to in the paper by Diesendruck & Markson (2001) confirm these results.

The first account on offer, as presented above, postulates the existence of a specifically lexical constraint, *mutual exclusivity*, compelling learning children towards inductive hypotheses that avoid this type of redundancy.

The competing account invokes a more general mechanism that is able to explain the same redundancy avoidance effects, as well as many others: children involve their mindreading abilities, their social and pragmatic understanding of the situation, and rationalize the speaker's behavior. They figure that avoiding lexical overlap is the most reasonable strategy for languages and speakers to conform to when the goal is communicative success, and that it would make less sense to go through the superfluous trouble of possessing a second word to do the same job of another word one already has. Accordingly, by default they assume that the novel name must refer to the novel object.

Diesendruck and Markson devised an empirical test to adjudicate between these two accounts. In Study 1, they reasoned that if the avoidance of lexical overlap was really due to rationalization, analogous effects should be manifest if, instead of *lexical items*, children were dealing with *associated facts*. The children were divided into two groups.

In *label-condition,* children from one of the groups are presented with two unfamiliar objects and taught a novel name for one of them - the experimenter picks up one of the objects and introduces the novel name, for instance, "Look at this one, it's a *zev*. See, it's a *zev*. This is a *zev*." Subsequently, they are asked for the referent of a second novel name, for instance, "Can you give me the *wug*?"

In *information-condition*, the children from the other group are presented with the same two unfamiliar objects but are taught a fact about one of the objects, not its name, "Oh, look at this one. My sister gave this to me. See, my sister gave this to me. My sister gave me this." Subsequently, they are asked for an object specified by means of a new fact, for instance, "Can you give me the one my dog likes to play with?"

Results show that children in label condition inferred the experimenter's intention to refer the nameless object just as naturally and promptly as children in the information condition inferred the experimenter's intention to refer the storyless object. This strongly

supports the pragmatic hypothesis over the lexical constraints account. In responding as they did, children were not conditioned by a mechanism specific to lexicon acquisition. This could not have influenced their responses in the second condition, but something did influence them. Instead, what determined their responses in both cases was something with a wider reach, a more general mechanism, their pragmatic understanding of the situation, rendering names and facts of equal use to them in allowing them to infer what really mattered - the speaker's referential intention. The child seems to have reasoned somehow along the following lines: if the experimenter had meant the first object, if she had wanted me to give her the first object, the rational thing for her to do would be to signal it using the name/piece of information we commonly know[98] to be associated with it. Since she did not, she must have intended to refer to the other object.

The mechanisms at stake are not specifically lexical, but general in their reach.

This in no way implies that children are blind to the specificities of words and linguistic information. Another experiment presented in the same paper shows that children expect linguistic knowledge to be widely shared across the community, unlike knowledge of other particular facts.

In the next two studies the child had two interlocutors instead of only one. In the first condition, the *knowledge-information condition*, the second interlocutor would be present, hearing, together with the child, the new information offered about the first object, a new name or a new fact. In the second condition, the *no-knowledge-information condition*, the second interlocutor would not be present to hear the new piece of information. In the test phase, the second interlocutor would ask the child for one of the objects using either a new label or a new piece of information.

What happened was that children would only not assume that the second interlocutor was asking for the second object in the *no-knowledge-information condition,* and when the information consisted of a fact and not a label. In such cases, when the unschooled interlocutor tried to ask for one of the objects by means of some new fact, children picked an object randomly. Contrastingly, when dealing with labels, children seemed to assume that the second interlocutor possessed the information conveyed to them by the first interlocutor, even

[98] There is common knowledge when both agents know something, and know that they both know it, and know that they both know that they both know it, and so on.

in the *no-knowledge-information condition*. When the second interlocutor asked them for an object by using a new label, they would assume she was not referring to the first object for which they already possessed a different label and so pick the other one.

This seems to reveal that they understood the association between label and object as a piece of linguistic knowledge and knew enough about language to assume that a competent speaker like the second interlocutor would most likely not fail to possess it as well.

4.3. CONCLUSION

Based on the experimental studies presented here, and others in the same line, theorists have come to almost form a consensus around the idea that mindreading is a crucial element in the process of word learning. After paying more close attention to such claims and elaborations, we can now confirm that the sort of inquiries children embark in into the minds of speakers and the communicative intentions behind their utterances are convincing instantiations of the type of rationalization of speakers and utterances that Davidson expected his interpreter to perform.

From the earliest stages of language learning, children exhibit an impressive understanding of other people and of the social environment and practices, one which continues to quickly grow and mature during their first few years. They assume in the other an equivalent perception of the world and of its objects and properties, and the coherent integration of their states and actions. As a result, they are able to recognize, often by unequivocal rational inference, their interlocutors' interests, beliefs, goals, requests, frustrations, changes of plan, meanings, references, and much more. These are essential resources in the process of becoming competent speakers.

Hence, it is appropriate to assert, in conclusion, that Davidson's foundational account of meaning and communication, as well as the thesis under scrutiny here that interpretation requires rationalization, enjoy ample support from the empirical study of word learning.

Part II

Alternatives to Rationalization

Introduction to Part II

In this second part, I will consider two alternative approaches to interpretation, two different ways of looking at the topic that promise, or at least suggest, ways of going about interpreting that do not involve rationalization.

The first consists of understanding linguistic abilities and practices as a natural phenomenon, to be studied and accounted for involving the methods of natural sciences. The proposal I will be considering, Quine's, also contemplates what we can describe as appropriately naturalistic methods of interpretation. I will deal with this Naturalistic Alternative in Chapters 5 and 6.

The second alternative emphasizes the conventional element in language and linguistic interchanges, and defends that regular instances of meaning ascription and interpretation are exclusively based on common knowledge of codified meanings. I will deal with this Conventionalist Alternative in Chapters 7 and 8.

Part II.a

Rationalization and Naturalization

Introduction to Part II.a

Part II.a is devoted to Quine's alternative approach to language, linguistic practices and interpretation, his naturalistic semantics, and comprises the critique of more traditional approaches, Chapter 5, and the first steps of a positive account, Chapter 6.

The two chapters could also be described as a critical presentation of Quine's epistemology - his conception of it, and an illustrative application of those ideas. For Quine, epistemology is the central philosophical discipline. Even his claims, positions, or interests that are not flagrantly epistemological in nature are often elaborations and developments of epistemological ones. This is the case with his study of meaning and language. Quine tackles the problem of *language, language acquisition and interpretation* as a problem of *theory and theory formation, learning and translation*. Accordingly, he identifies his foundational semantics - his attempt to identify and understand the underlying facts that constitute what emerges as an apparent semantic reality - with his epistemology - his attempt to understand the processes through which humans attain knowledge, i.e. arrive at their theories of the world.

In Chapter 5, I start by discussing the theoretical basis of Quine's program. His naturalism motivates a radical departure from tradition in his conception of epistemology. Rejecting the traditional normative type of investigation, Quine defends that the study of knowledge must take the form of scientific psychology, not so much worried about the rational credentials of the epistemic states but aiming instead at a faithful description and understanding of the actual processes through which they come about. On my part, I resist the idea of a complete overcoming of the old style of inquiry. I note that there is a component of traditional epistemology in the interpreter's rationalizing approach, and defend its likely indispensability.

Quine's naturalism incorporates a particular conception of science, and of genuine explanatory progress, that determines the pursuit of the most elementary-level descriptions and accounts of the object of study, in detriment to, in particular, mentalistic and rationalistic explanations. In the study of language he rejects the mental level approach, and defends that

any real elucidation can only start at the behavioral level and be continued at neural or physiological - and perhaps even chemical and physical - levels. What is implied here is the lack of relevance and theoretical value of the descriptions and explanations presented and endorsed in Part I of this dissertation. In response, still in Chapter 5, I note the thinness of the grounds on which Quine bases his ranking of sciences and, especially, his animosity towards the mental.

However, the best prospect for a Quinean argument in favor of the need for the overcoming of a mentalistic approach to interpretation and language acquisition will be fully explored only in the following chapter, Chapter 6. It consists of attempting to prove the superfluity of the mentalistic account by means of offering an alternative competent account of the same phenomena in non-mentalistic terms. Quine sketches such an account. I defend the position that his rudimentary theory offers an inadequate characterization and understanding of language and linguistic practices and, hence, that he completely misses this opportunity to advance his case against the worthiness of mental idioms and explanations.

In the second part of Chapter 6, I distinguish and examine three different plans for interpretation that can be found in Quine's work. Two of them comport an apparent threat to the principal thesis of this dissertation - that interpretation requires rationalization - as they pose as naturalistic alternatives to the methodology defined in detail in Part I. I defend the thesis that the threat is merely apparent and that there is no credible alternative method of interpretation, one genuinely dispensing rationalization, to be found or guessed in Quine.

5. Naturalistic Rejection of Rationalization

In considering Quine's naturalistic alternative account of language and linguistic practices, I start with the reasons and arguments he uses to support his original departure from more obvious paths, or environments, of investigation. Quine attempts to settle his accounts and explanations as far as possible from our common-sense, pre-theoretical, mentalistic conceptions on these issues.

There are two ideas I render particularly salient in my examination because, besides motivating and preparing the way for Quine's own positive proposal - to be expounded in the next chapter - they are also in direct conflict with the rationalizing interpreter's methodology. These are the rejection of the mentalistic idiom and explanations, and the abandonment of a normative and internal perspective on epistemological inquiry. Both points are placed in context as consequences, or realizations, of Quine's more general and fundamental positions: Quine's naturalism, together with a very strict conception of science, and a primary interest in epistemology. I inspect more closely each challenge and mount my resistance.

I start with Quine's anti-mentalism. I try to elucidate Quine's reasons and grounds for the rejection but find them lacking or unconvincing. There is no general idea of a scientific method sufficiently worked out to support Quine's rejection. The flaws and insufficiencies he identifies in such idioms and accounts do not seem to prevent them from being very useful in diverse fields of study. I note that very successful disciplines employ psychological explanations and/or make essential appeals to mental entities or processes.

Next I turn to Quine's rejection of traditional epistemological methods and concerns. I present an important objection frequently pressed against his revisionistic proposal, the *Normativity Objection*, and consider Quine's response to it. I defend that, independently of more theoretical concerns regarding the fate of the epistemological *discipline*, there is an epistemology of an intuitive and natural type ingrained in our basic perspective over our own and other people's mental states. More, I note that this epistemology is an essential element in interpretation and language acquisition.

5.1. EPISTEMOLOGY AND NATURALISM

Quine gets to language and the study of language through a more general interest in knowledge and the study of knowledge. Accordingly, we will also start with his epistemology. Quine's epistemology is of interest to our topic not only because, in a general and indirect sense, epistemology concerns all particular disciplines, scientific enterprises and knowledge progresses, with any science of language and the very feat of language acquisition, of course, included. Besides that, there is also an even more direct connection between epistemology and the study and reflection on language and language acquisition and interpretation. Quine thinks of language as embodying knowledge, hence, for Quine, to a large extent, the study of knowledge and of knowledge acquisition coincides with the study of language and of language acquisition. As he says: "epistemology becomes semantics" (Quine, 1969a, 89).

Arguably, the most salient aspect in Quine's philosophical path is his commitment to naturalism. Quine´s naturalism takes form, first of all, in a strong conviction that all human knowledge[99], in all its aspects, constitutes one continuous, integrated, single affair, with no real drastic separations within. According to Quine, we should beware of any radical splits inside the epistemic domain, be it between philosophy and science (Quine, 1969), between common sense and science (Quine, 1957), or between the analytic and the synthetic (Quine, 1951). They are either false as claims, or pernicious as inquiry strategies. There are no clear and distinct separations, no perfectly autonomous niches, no strict and irreversible hierarchies or priorities. All attempts and all real progresses contribute to a common edifice. Some of the actual contributions are better than others; some of the actual practices deliver better results than others. The best we can aspire to is science - what Quine counts as science is a good question to which we will be returning afterwards - but, of course, not all occasions require us to meet or try to meet the best scientific standards.

Quine's naturalism strongly determines his epistemology. We must distinguish two components in Quine's views on the subject. A two-pronged negative and critical one that combines the rejection of traditional epistemology and of intentional discourse and explanations grounded in our folk understanding of the mental; and a positive one, a proposal about how to move beyond the traditional program by converting epistemology into science,

[99] He has strong reservations concerning the notion of knowledge but, for our purposes, it will do.

and doubtful explanations into genuine ones. I will start with Quine's critique of traditional epistemological approaches.

5.1.1. THE ABANDONMENT OF TRADITIONAL EPISTEMOLOGY

Quine identifies a '*duality of structure'* (Quine, 1969, 71) in traditional epistemological approaches. On the one hand, we find the kind of concerns he refers to as '*conceptual'* (ibid., 69), while on the other hand, the ones he refers to as '*doctrinal'* (ibid., 69).

On the conceptual side of traditional epistemology we find problems related with the meaning and content of our beliefs. One is laboring on the conceptual side, for instance, whenever one is trying to define less fundamental concepts in terms of more fundamental ones. To exemplify the sort of ambition that fits here, Quine refers to the canonical attempts of translating mathematical terms into logical terms, and of translating talk of bodies in terms of sense experience.

On the doctrinal side, we find problems related with truth, certainty and proof. On this side of the enterprise, the general goal is that of grounding beliefs and scientific claims. We have, for instance, the project of proving the truths of mathematics, by deduction, from self-evident logical truths, or the classical attempts to ground one's knowledge of nature in supposedly indubitable truths concerning one's phenomenological experience.

In none of these components did the traditional epistemological project achieve success. The best attempts in each field have all failed and been abandoned. Just as, on the doctrinal side, "the hopelessness of grounding natural science upon immediate experience in a firmly logical way was acknowledged" (ibid., 74) so, on the conceptual side, "philosophers have rightly despaired of translating everything into observational and logico-mathematical terms" (ibid., 82). What should we think and how should we continue epistemology having recognized this discouraging state of affairs?

In order to get ready to answer this question, let us consider first the place of science - or maybe the lack of it - in traditional epistemology. In Quine's diagnosis, if traditional epistemology is not free to make use of science and of its products and revelations in pursuit of its goals, that is because traditional epistemology still rests on an idealized and unviable conception of knowledge. It is because it aspires to a certain idea of cognitive perfection that

it cannot employ and rely on science, as necessary shortcomings would follow on both the conceptual and the doctrinal projects. Reliance on science would not allow us to go beyond a circular justification of our knowledge. Such a justification of our knowledge would be grounded on the very presupposition of its truth, thus frustrating the foundational ambitions that animate the doctrinal side. On the conceptual side of the challenge, any scientific explanation would again fall short of a satisfactory answer. Science cannot be used to elucidate the concepts, contents and judgments that we care to elucidate because the concepts, contents and judgments that we care to elucidate are, first of all, those very same scientific concepts and judgments that we must already understand and employ in order to produce any type of scientific account. Science itself employs concepts and produces statements that far exceed the mere report of input - whatever that is - and thus must already presuppose their intelligibility before it can explain anything.

It is time to return to the question in the penultimate paragraph, as we can now provide an answer. What we can and must do after accepting the chronic shortcomings of traditional epistemology, and the consequent imperfection of our knowledge, is to recognize the new liberty resulting from it - that of employing science in epistemological investigation - and to effectively explore it.

5.1.2. NATURALIZED EPISTEMOLOGY

On the one hand, having relinquished the ambition of doctrinal reduction, we have no other reason to refuse the mobilization of science - in other words, "scruples against circularity have little point once we have stopped dreaming of deducing science from observations" (ibid., 76). On the other hand, having despaired of any complete conceptual reduction, "a true reduction by translation, a legitimation by elimination" (ibid., 78), the second best thing to hope for from a rational reconstruction, in terms of conceptual clarification, presents no advantage over the kind of account provided by science itself. And so:

> If all we hope for is a reconstruction that links science to experience in explicit ways short of translation, then it would seem more sensible to settle for psychology. Better to discover how science is in fact developed and learned than to fabricate a fictitious structure to a similar effect. (ibid., 78)

Quine exhorts us now to the naturalistic, scientific and empirical study of the actual

processes through which the formation of knowledge and the building of science take place. Quine recommends that, henceforth, this is what the epistemological inquiry must consist of. This is the positive and constructive component of his views on epistemology.

> I think that at this point it may be more useful to say rather that epistemology still goes on, though in a new setting and a clarified status. Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. This human subject is accorded a certain experimentally controlled input - certain patterns of irradiation in assorted frequencies, for instance - and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. (ibid., 83)

The emphasis in such an epistemology is on the external, third-person description of the processes through which knowledge is formed. Any possibility of critical assessment of that knowledge can no longer be a radical one. Instead, such a possibility is now seen as subordinated to a previous and general acceptance of our knowledge and science as our starting point, to a large extent good and worthy, but always admitting criticism, revision and improvement - the reconstruction or repairment of a boat while afloat is Neurath's celebrated metaphor for that task. Naturalized epistemology is hardly interested in justification, let alone independent justification. In naturalized epistemology, one starts by assuming that there is knowledge, and then one explores that initial capital and tries to make it grow, albeit in a particular direction - that of elucidating the very processes by which such knowledge has come about.

This project of naturalized epistemology sets the stage for the rest of the discussion in this chapter. What follows is the search for a clearer understanding of the conception of science involved, and what Quine is rejecting with his very strict commitment to it.

5.2. CONTROVERSIES WITH QUINE'S NATURALISM

Many of the arguments, ideas and stances that Quine mobilizes in the defense and promotion of his naturalism and naturalized epistemology - the abandonment of foundationalist ambitions, the abandonment of the project of reducing or translating physics to phenomenology, the acknowledgement of the relevance of natural sciences in the study of knowledge, belief, perception and so forth, among others - enjoy broad consensus. Yet not

many authors seem to end up fully convinced of the pertinence of a complete and radical shift of the epistemological paradigm in the terms and fashion determined by Quine. To be sure, Quine's proposal had a tremendous impact at the time, and gave a great new breath to a naturalistic strand in epistemology that already existed before but bloomed only afterwards. But it gave rise to as many and various forms of endorsement as to forms of rejection.

I distinguish two families of problems, two lines of criticism, that are of special interest to my study. The first has to do with Quine's peculiar conception of science. There might be reasons for doubts and perplexity over Quine's ranking of sciences, in particular about his disregard for all disciplines and inquiries with some close relation to the mind and the mental. In the second group, we have the problems that constitute or relate to what is usually referred to as *'the normativity objection'* to naturalized epistemology. Both groups of problems are, of course, related.

5.2.1 QUINE'S CONCEPTION OF SCIENCE

As we have seen, with a specific incidence on epistemology, Quine's naturalism consists of a certain form of scientism: in short, the belief that there is no knowledge outside science, that philosophical and epistemological theories and proposals must be pursued employing the same methods, and they are to be assessed according to the same standards.

It is time to return to the first question left suspended a few pages above: What is Quine talking about when he talks about science? Quine's conception of science determines his epistemology, which, in turn, determines his view of language and language acquisition and of what the study of language can and should be. In order to get a clear understanding of any of the latter, one needs to have a more precise and elaborate understanding of the first.

Here is an early passage from Quine's work that is very elucidative of the kind of descriptions and accounts of processes and events that the naturalized epistemologist, according to Quine, should be interested in.

> I am a physical object sitting in a physical world. Some of the forces of this physical world impinge on my surface. Light rays strike my retinas; molecules bombard my eardrums and finger-tips; I strike back, emanating concentric air waves. These waves take the form of a torrent of discourse about tables, people, molecules, light rays, retinas, air waves, prime numbers, infinite classes, joy and sorrow, good and evil.

> My ability to strike back in this elaborate way consists in my having assimilated a good part of the culture of my community, and perhaps modified and elaborated it a bit on my own account. All this training consisted in turn of an impinging of physical forces, largely other people's utterances, upon my surface, and of gradual changes in my own constitution consequent upon these physical forces. All I am or ever hope to be is due to irritations of my surface, together with such latent tendencies to response as may have been present in my original germ plasm. (Quine, 1957, 215)

What is striking is the elementary level at which Quine suggests the descriptions can take place. Speaker, thinkers, persons, appear depicted as physical objects, tremendously complex to be sure, but destitute of mind and intelligence, inadvertently reacting to the forces, being causally shaped into new physically describable states that, in some sense, correspond to those of possessing a language and a theory.

Quine admits different levels and kinds of naturalistic accounts. First in Quine's ranking is physics but, of course, he also allows the naturalized epistemologist other approaches. He frequently mentions biology, neuroscience and behavioristic psychology.

These, however, do not exhaust science. On the one hand, there is also logic and mathematics. Despite Quine's rejection of the analytic-synthetic divide, he still recognizes the distinctive character and pervasive role of these disciplines. Quine also acknowledges the "softer sciences, from psychology and economics through sociology to history" (Quine, 1995, 49), but he is very clear in not seeing them on an equal footing with his preferred hard sciences. As Gary Kemp aptly puts it, he "posits a continuum of rigour, objectivity, and explanatory potential, with mathematical physics at one end, and history, grading off into journalism, common sense, and gossip, at the other" (Kemp, 2012, 16). What is much less clear is what reasons lie behind this ranking, and how good they really are.

Chomsky maintained a long dialogue with Quine over the years and famously criticized his naturalism and conception of science - as one can say, *from within the natural sciences*. Here he expresses disagreement and a certain puzzlement about the substance and grounds of Quine's choices.

> Take the version of this doctrine expressed by Quine ... In his most recent formulation, the "naturalistic thesis" is that "the world is as natural science says it is, insofar as natural science is right." What is "natural science"? Quine's total answer is: "theories of quarks and the like." What counts as like enough? There are hints at answers but they seem completely arbitrary, at least by ordinary naturalistic criteria. (Chomsky, 2000, 144)

Along a similar line, Alan Wier, reflecting on this issue in a recent text for a collected volume on Quine's philosophy (Wier, 2014), makes a very compelling exposition of the obscurity and dubiousness of the criteria used by Quine to draw the line between what counts as worthy science or explanation, and what does not.

I think Wier is right when he defends that Quine's judgments about the scientific credentials of disciplines and explanations can hardly be accounted for in terms of a better or worse conformity of those disciplines and explanations to some scientific method or standard. What Quine has to say about a general method for science is very little and arguably not enough to support the kind of ranking he proposes. Quine's allegiance to the hypothetico-deductive method[100] goes together with a few regulative principles favoring simplicity, conservatism, ontological parsimony[101], and assorted remarks promoting intersubjectivity, objectivity, empirical testing, precision, the integration of sciences[102], and maybe a few other things. These are somewhat vague terms and principles, and do not seem to provide enough guidance and assurance in delicate issues such as the discrediting or downgrading of notions or of established scientific practices with clear positive results. Besides, even such principles and maxims are hardly justified themselves; they can be challenged, or they can be defended to be valid or useful only relatively to specific interests.

> [H]ere the clash between "hermeneuticists" and "hard-line naturalists" comes to the fore. The hermeneuticist may acknowledge the astonishing successes of physics and chemistry as validating the methodology of those austere sciences in their own domains but deny that this provides any grounds at all for supposing the same methodology will bear fruit when applied to the social sciences. When trying to make sense of rational beings who act, individually and collectively, on the basis of beliefs, intentions, and goals, our hermeneuticist says that empathy and an ability to get into the mindset of the objects of study, not hypothetico-deductive method, are the order of the day. (Wier, 2014, 121)

We start to see a little more clearly how it is that Quine's endorsement and interpretation of naturalism falls short of consensus - to mark its being a particular conception

[100] Professed, for instance, here: "Such generalization, called simple induction, distinguishes us *none* from other high mammals except that we verbalize it. It only puts words to their habitual expectations. Sophisticated science far transcends simple induction. It interpolates unseen interim careers of seen things (see THINGS) and fabricates terms for fancied things unseen. It posits abstract objects, notably numbers, and with help of these, devises measurement. Thus arises a powerful and virtually conclusive refinement of induction, the method of which John Stuart Mill called concomitant variation. An intricate web of hypotheses is devised which together imply a host of observation categoricals. Such is the "hypothetico-deductive method"." (Quine, 1987, 161)
[101] See, for instance, Quine, 1995, p.49.
[102] All these ideas are already present, for instance, in Quine, 1957.

or interpretation of naturalism, from now on I will be referring to it as *(Quinean-)naturalism*. It is not certain that his naturalistic and scientific principles are the best suited, nor that they determine the hierarchy of disciplines he defends, let alone that these principles cannot co-exist with other, eventually non-naturalistic, principles associated with different interests and different epistemic domains.

There is another point in Wier's argument that I want to echo. I also agree with him when he claims that it is not the case that the accounts or disciplines Quine urges us to abandon have been demonstrated by him to be *'explanatory redundant'* (Wier, 115). Quine has not come close to start explaining how history, or anthropology, or decision theory, just to mention a few, can be substituted or supplanted by some different account in Quine's preferred vocabulary. In Chapter 6 I hope to partially vindicate this claim by showing the inviability of Quine's sketched proposal for a naturalistic account of language, communication and understanding.

5.2.1.1. QUINE'S REJECTION OF MENTAL EXPLANATIONS

One of the most prominent features of Quine's naturalism and peculiar conception of science is his profound suspicion concerning all the disciplines and epistemic endeavors that one way or another refer to minds and mental states and employ mentalistic discourse.

> My position is that the notions of *thought* and *belief* are (...) ill suited for use as instruments of philosophical and scientific clarification and analysis. (Quine,1981b, 184).

Along the years, Quine repeats his conviction that appeals to intentionality or to the mind will mainly tend to obstruct the possibility of effectively elucidative accounts. Being too obscure and imprecise, according to his view, these notions are unfit for science and unable to generate genuine progress in our understanding of the subjects. Worse, the use of such notions can only bring about the illusion of explanations and, that way, present the danger of hindering serious science.[103]

[103] He expresses it, for instance, here: "People persist . . . in talking of knowing the meaning and of sameness of meaning. . . . They do so because the notion of meaning is felt somehow to explain the understanding and equivalence of expressions. We understand expressions by knowing or grasping their meanings; and one expression serves as a translation or paraphrase of another because they mean the same. It is of course spurious explanation, mentalistic explanation at its worst. . . . where the real threat lies, in talking of meaning, is in the illusion of explanation." (Quine, 1975b, 86–87)

In response, a number of critics have defended the opposite, i.e. that mental objects, notions and accounts are neither particularly obscure and imprecise, nor destitute of explanatory power. I agree with these critics, and I will briefly present the case of some of them.

Quine continues the passage above and completes his thought explaining that "If someone accepts these notions outright for such use" (ibid) he is "at a loss to imagine what he can have deemed more in need of clarification and analysis than the things he has thus accepted". (ibid.) It seems, however, that it is not hard to answer Quine's perplexity. As much in prosaic everyday contexts as in scientific ones - both in human and natural sciences - one finds plenty of examples of useful and elucidative accounts employing these and other mental notions.

Starting with the prosaic ones, one surely *'can have deemed more in need of clarification and analysis'*. For instance, why has Charles walked to the fridge, shot the gun, crossed the road, or uttered 'I will meet you there at eight'. Plausibly, if one is troubled by any of these questions, or similar ones, one will find no good, satisfying answer outside intentional discourse. This is not in contradiction with Quine's remark. There he rejected its usefulness in science and philosophy, not in mundane affairs. It is just as easy, however, to find equally good examples within science.

In a general commentary and critique to Quine's repudiation of the mental, Alan Wier illustrates this point with history, noticing that "[t]he historian explains the Allied decision to invade Normandy as partly the result of a belief that Hitler did not expect the Allies to invade there". (Wier, 2014, 139). Wier puts forward a number of disciplines where, again, theorists routinely 'deem' their particular questions and subject-matters "more in need of clarification and analysis" than the repudiated notions they happen to employ on their account. He lists 'synonymy in lexicography' (ibid., 136), 'beliefs, desires, aims, or goals in history or anthropology' (ibid.), 'content, proposition and belief in cognitive psychology or decision theory' and 'representation in computational theories of perception' (ibid).

This critique comes also from science. Chomsky sees himself as a natural scientist - with respect to his work on language - and strongly defends his case[104]. Famously, he elects the mind and the mental as an object of study, and appeals to mental structures and

[104] See Chomsky, 2000; see also McGilvray, 2005.

mechanisms, although not those of folk psychology. Quine opposes this, both explicitly, directly addressing Chomsky and his reference to the mind in the form, for instance, of implicit rules (see Quine, 1970), and indirectly, via his intransigent assertion of opposing views, in particular those concerning language acquisition and *the indeterminacy of translation* (see, for instance, Quine, 1960, and Quine, 1995). Chomsky accuses Quine and his stripe of naturalism of doing the opposite they profess, of disregarding science and of colliding with genuinely naturalistic methodologies. Here, for instance, he criticizes Quine for his blindness to empirical evidence and for what he takes to be a groundless rejection of the explanatory potential of posits such as the Language Acquisition Device (LAD),

> The empirical evidence is overwhelming that association and conditioning have little to do with language acquisition or use, but that seems not to matter; one wonders why. Whatever the answer, we find examples of what Quine favors (quarks, neural inputs, conditioning) and disfavors (the devices of LAD, that is, the operative mechanisms, so far as is known). But we are offered no reasons for the decisions, or more than a few examples to suggest their scope. (Chomsky, 1994, 92)

For yet another critical voice, take Tyler Burge - interestingly, a student of both "Quine (in absentia) and Davidson (viva voce)" (Burge, 1999, 230). Burge's special proficiency and attention to science - in particular, psychology, biology, and neuroscience applied to the study of perception and cognition - and the way in which that science actually informs and determines his philosophical positions seem to position him clearly in a naturalistic lineage. However, he does not share Quine's scruples against the mind and the mental. In particular, in his *Origins of Objectivity* (2010) Burge presents a robust defense and exemplification of the importance and aptness of the mentalistic notions for the sciences of perception. For instance, already in the preface, he explains,

> I emphasize mind not because I think that minds float free of brains, or other aspects of physical reality. I think quite the contrary. I emphasize it because I think that explanations and descriptions in mentalistic or psychological terms provide deep, scientifically indispensable insight into the way things are.
>
> …
>
> Science itself - most impressively vision science, but more broadly perceptual psychology and developmental psychology - has vindicated psychological, mentalistic notions. The explanatory power of the sciences vindicates these notions' viability for scientific purposes. The emergence of mathematically and explanatorily rigorous explanations in perceptual psychology, and the use of results from perceptual psychology by sciences like animal psychology and developmental child psychology,

place skepticism, hostility, patronization, and unease about the scientific value of psychological notions at odds with science itself. (Burge, 2010, xiv)

The idea, formulated in this passage and advocated by Burge throughout this book, of mind and mental idioms playing an indispensable role in the science of perception is especially relevant to us here. I will show later in the text how the success of Quine's project of a *(Quinean-)naturalistic* account of language is dependent upon the possibility of an equally (Quinean-)naturalistic account of perception and of perceptual classification. Thus, if Burge is right, Quine's project will be compromised. Quine, on his part, does not himself have any mental-free explanation of perception to offer - I will stress this point later in the next chapter. Notably, he seems to have no particular idea about how *perceptual similarity* can be (Quinean-)naturalistically realized and understood. In fact, to hold that aspects of a human's cognitive life such as perception and perceptual classification should or could be studied without the employment of mental discourse and notions, reference to mental entities, mechanisms or events, or uncomplemented by other scientific enterprises that do use and refer to them, seems to be a minority position nowadays.[105]

The previous examples are more than enough to make it clear that Quine faces substantial disagreement concerning his negative opinions about the explanatory and scientific value of the mind and of mentalistic idioms. The inutility or dispensability of the mental is far from obvious. Eliminativist positions of the kind that Quine's fiercest passages against the mind seem to prescribe are rare, at least these days. I will not delve deeper into any of these particular cases to try to settle the issue and determine who is right in each, nor will I defend a general position on the topic of eliminativism. Instead I will bring my focus once again to Davidson and his interpreter, and to the empirical disciplines explored in Chapter 4. I limit my defense of the explanatory merits of mental accounts to the topics of interpretation and of language acquisition and the learning of word meanings. We saw how Davidson depicts and explores the psychology of the process to explain and make sense of interpreters' successes, and we saw how congenial accounts are defended in recent scientific research and literature on lexicon acquisition.

In Chapter 4 we considered three competing views, or accounts, of word learning and

[105] See, for instance, Ramsey, 2016, Hardcastle, 2009, McCauley 2007 and Paul Thagard, 2013.

a number of experimental studies designed to arbitrate between them. To start with, those are views and studies in psychology and, hence, views and studies targeting the mind. In that sense, we are wholly and firmly in the mental domain, but this is not enough for our point. Quine's reluctance is not so much with taking the mind as an object or target of scientific clarification, but rather with taking the mind - and appeals to its processes, operations, mechanism, structures, events, objects - as an actual means of such clarification. However, the latter too is the case with the theories considered. *Associations*, *lexical constraints*, *minds* and *mindreading skills* are flagrantly mental postulates and they are employed, with success, in attempts to explain something else, the acquisition of the lexicon. Moreover, among the psychological structures appealed to, we also find states and processes of the most commonsensical sort - beliefs, desires and intentions, in particular, communicative intentions. They are accepted in the field and they are doing some actual work in the best proposals available.

As for Davidson, he in no way shares Quine's rejection of the mental. He openly adopts mentalistic accounts and believes that some illumination comes from this, even if of a different kind, and employing different standards of rigor, regularity and clarity from those one would expect from an ideal nomological science. *Hold true* in Radical Interpretation is not a conditioned reaction but a rational action or state on whose account one has to appeal to mental states. As an action, hold true is explainable as the result of a certain desire and of a certain belief about the way of satisfying it. As an action of a particular type - a typical affirmative utterance from a cooperative speaker - it is explainable, for instance, as the product of the agent's belief that the sentence in question has a certain meaning, together with another belief with the same content as the sentence held true, and with some desire to express that belief.

Davidson does not expect a (Quinean-)naturalistic account of language use, acquisition or interpretation to be possible. That is, he believes the relevant aspects of linguistic phenomena are only graspable and explainable with intentional vocabulary. Davidson is convinced of *the anomalism of the mental*. In its barest form, this is the claim that "there are no strict laws on the basis of which mental events can be predicted and explained". (Davidson, 1970, 208). *Mental events* include all events describable in the intentional idiom, such as *holding true*, *believing that*, *intending to* and still other propositional attitudes essential for a rational and rationalizing approach to language

acquisition or interpretation. *Strict laws* are the kind of laws that constitute the core of the kind of naturalistic account ultimately envisaged by Quine. If the vocabulary that Davidson believes able to describe the relevant aspects of language and communication and allow an agent to find her way into proficiency in these practices cannot be translated into that of the hard sciences, then he cannot hope the naturalistic account Quine seeks will be adequate. That is, Davidson's commitment to the anomalism of the mental implies a radical disagreement with Quine's naturalistic project.

This, in turn, suggests a certain strategy for the defense of the claim I am endorsing here - the claim that rationality and rationalization play an indispensable role in language acquisition and interpretation. If anomalism is right, and if anomalism implies the inadequacy of a naturalistic approach to language, i.e. its inability to recover the features that matter to reconstruct the linguistic communication as we recognize it, then a rational and rationalizing approach, being the only alternative left, could claim for itself the task of illuminating the subject. But that is not a strategy I will follow - not, at least, at such a general and abstract level. Instead I choose a more detailed and positive line of defense. I want to show in the next chapter how Quine's particular naturalistic approach is unsatisfying, in contrast with a rational and rationalizing account that, as we saw in Part I, seems indeed to offer some fruits, some predictive and explanatory value, plenty of practical gains, and to be employed and enjoy wide recognition not only in everyday life but also in science, in psychology and developmental studies

5.2.1.1.1. AMBIVALENCE AND TENSION

The reference to the anomalism of the mental at the close of the preceding section gives us occasion to elaborate a little on Quine's conception of the mind and mental, adding some nuance and complexity - as well as some tension - to what I have said so far.

First of all, I must not omit that Quine himself, at times, admits some explanatory potential to mentalistic accounts. Already in *Word and Object*, we find him declaring that he "would [not] forswear daily use of intentional idioms, or maintain that they are practically dispensable" (Quine, 1960, 202), and, at least in later writings, he even acknowledges some vestigial scientific value in them.

> … the mentalistic predicates, for all their vagueness, have long interacted with one another,

> engendering age-old strategies for predicting and explaining human action. They complement natural science ... and are indispensable both to the social sciences and to our everyday dealings. (Quine, 1990, 73)

In the face of passages such as this, we must, perhaps, adjust the previous anti-mentalistic picture. The sensible thing to do, it appears, is to take Quine's occasional harshness towards the mental as, in part, rhetorical exuberance, most plausibly directed at stressing that, even if mental accounts can have merits of their own, his interests lie elsewhere, with other sorts of descriptions. On this last point, he is unrepentant and perfectly unambiguous throughout the years: he believes it possible, and seeks to contribute to the progress of a naturalistic account of human knowledge and language acquisition.

This reading is consistent with certain other passages where Quine subordinates the choice of vocabulary to the goals and interests in question.

> [T]hey call, I think, for bifurcation in canonical notation. Which turning to take depends on which of the various purposes of a canonical notation happens to be motivating us at the time. If we are limning the true and ultimate structure of reality, the canonical scheme for us is the austere scheme that knows no quotation but direct quotation and no propositional attitudes but only the physical constitution and behavior of organisms. … If we are venturing to formulate the fundamental laws of a branch of science, however tentatively, this austere idiom is again likely to be the one that suits. But if our use of canonical notation is meant only to dissolve verbal perplexities or facilitate logical deductions, we are often well advised to tolerate the idioms of propositional attitude. (Quine, 1960, 202-3)

It would be very convenient to be able to understand these choices and *bifurcation* as a merely pragmatic issue: it would involve seeing the employment of the mentalistic idiom as better suited for certain less rigorous tasks and disciplines - facilitating the understanding and report of certain (broader and synoptical) aspects of things - while, at the same time, preserving the ability to transit to a (Quinean-)naturalistic and austere vocabulary, so as to increment the detail and rigor of the enterprise, if the occasion should justify it. However, we cannot hold on to such an interpretation, as Quine is not in a position to countenance such a reconciliation. Quine has, for a long time, endorsed some form of the anomalism of the mental, and that precludes any chance of smooth transition between realms and discourses.

We find early manifestations, for instance in his adhesion to "Brentano's thesis of the irreducibility of intentional idioms" (Quine, 1960, 202), and also later ones, influenced as well by Davidson's pronouncements.

> The notion of an exhaustive class of states each of which qualifies as thinking about Fermat's Last Theorem, and each of which is specifiable in purely physiological terms, seems discouragingly unrealistic even if restricted to a single thinker. It is at this point that we must perhaps acquiesce in the psychophysical dualism of predicates, though clinging to our effortless monism of substance. It is what Davidson has called anomalous monism. Each occurrence of a mental state is still, we insist, an occurrence of a physical state of a body, but the groupings of these occurrences under mentalistic predicates are largely untranslatable into physiological terms. There is token identity, to give it the jargon, but type diversity. (Quine, 1995a, 87)[106]

Quine's belief in the anomalism of the mental, particularly when seen together with his acknowledgment of the indispensability of the mentalistic discourse, is in flagrant tension with his faith and commitment with naturalism and scientism, and with what he still retains of an ideal of scientific continuity and unity[107].

It might be tempting to read Quine's distinction between *real* and *spurious* propositional attitudes as pointing in the direction of a certain type of compromise. Here are three passages where the distinction is articulated:

> Some beliefs, perhaps belief in the essential nobility of man qua man, are indeed not readily distinguishable from mere lip service, and in such cases there is no fact of the matter by any reasonable standard. But most attributions or confessions of belief do make sense, within varying limits of vagueness. The states of belief, where real, are dispositions to behavior, and so, again, state of nerves. (Quine, 1986a, 429)

> ...ascriptions of perception get more tenuous and conjectural as we move from observational content clauses to others and finally to standing ones. Ascriptions of belief run more tenuous still, and their supporting evidence is diffuse. The ascription of beliefs by content clauses is disarming in its syntax: any declarative sentence is grammatically admissible as content clause, and hence is presumed to yield an intelligible ascription, however remote from any conceivable evidence. A belief can be as firmly testified in behavior as a tail-wagging dog's belief that his dinner is forthcoming. But what are we to say of a belief in the transubstantiation of the eucharist?
>
> The construction 'perceives that p' was essential to the propagation of language, and at the observational level it was well under the control of empirical evidence. By extrapolation, analogy, and further extrapolation, however, it has spawned a boundless lawless swarm: the ascriptions of belief. Responsible ones grade off into the irresponsible, and one despairs of drawing a line. (Quine, 1990, 67)

---

[106] This is not something Quine came to recognize only later, and any inconsistencies that may be found can hardly be attributed to a change of mind. Quine was writing about it already in *Word and Object*, 1960, 200-3. See also Quine, 1990, 71-3.

[107] See Quine, 1960, 231, and 1995b, 260.

> Perceptions are neural realities, and so are the individual instances of beliefs and other propositional attitudes insofar as these do not fade out into irreality altogether. (ibid., 71)[108]

Gary Kemp takes this line. In a recent book on Quine and Davidson (Kemp, 2012), right after presenting Quine's endorsement of the anomalism of the mental, Kemp raises the issue of Quine's ambivalence towards the factuality of the intentional idiom. He asks, rhetorically: "What better guarantee could there be for the reality of a given kind of discourse than its indispensability?" (Kemp, 2012, 63), and identifies the seeming instability in Quine's position, i.e. that he "seems to admit the idiom with one hand whilst taking it away with the other". Then he offers his remedy. He proposes that "the key to relieving" the discomfort with this apparent tension is to note "that some uses of the idiom convey facts, others do not; and the facts they do convey are more or less vaguely defined clusters of facts about dispositions" (ibid.). After quoting the same Quinean passage I have also quoted above - the first passage in the group of three - he goes on to explain the point in more detail.

> The looseness of most belief-ascriptions, such as '--- believes that Paris is in France', is endemic, but does not get in the way for ordinary purposes. Most ascriptions of propositional attitude indicate more-or-less loose concatenations of linguistic dispositions, which in turn are constituted by more-or-less loose concatenations of neural states. Thus, most of them do communicate factual information that is in principle naturalistically specifiable; however, the existence of those cases makes possible the formulation of others which do not. So the key is not to either ban or to accept the ... propositional attitude idioms as a whole, but to accept most within limits, while rejecting some. (Kemp, 2012, 63)

I disagree. I do not think we can take much relief from these observations. A distinction between more and less *real* propositional attitudes can in no way attenuate the tension between Quine's ideal of naturalistic hegemony and his acceptance of anomalism. The *real* attitudes would be just as untranslatable, just as naturalistically unaccountable as the others. An internal distinction within the intentional cannot help us here. If this does not result so clearly at times, it must be because the true nature and challenge of anomalism is

[108] I quote extensively because the point is problematic. These passages do seem to suggest something close to the interpretation proposed by Kemp that I will turn to presently. However, the view attributed by Kemp is incoherent with another view of Quine, his belief in anomalism, or the irreducibility of the mental - Kemp evades this problem. Since there is still room for doubt concerning Quine's actual point, I abstain from an assertive uncharitable reading.

misrepresented or watered-down. This is the case with Kemp's passage quoted above. He still suggests some type-type identity, even if a *'more or less vaguely defined'* one, between robust attitudes and something naturalistically specifiable, but that is precisely what anomalism rejects.

The apparent tension - between anomalism, naturalism and the tolerance of some mentalism - cannot be solved by appealing to the fact that Quine is willing to admit some revision of the explanandum. Trimming the set of propositional attitudes to exclude dubious ones such as *'the believe in the essential nobility of man'* does nothing to secure the possibility of a (Quinean-)naturalistic account of more *respectable* attitudes, if Quine is to hold on to anomalism and untranslatability.

There are some regularities that we can account for using mental idioms, which are the regularities that we care to see explained - for instance, the regularity of *holding true* the sentence '*There is a cat on the mat'* when one *believes* that the cat is on the mat. The untranslatability of these types of events into neurological or physiological ones means nothing less than the possibility of such regularities not being detectable, let alone explainable, at those different levels. With anomalism and untranslatability on board, going (Quinean-)naturalistic amounts to more than the ordinary scientific improvement or variation over some previous commonsense notion; it is closer to a complete change of topic. What would such a theory be explaining if not those initial regularities that we cared for?

Quine's disparate views and pronouncements on mentalism and the mind seem impossible to reconcile. The internal tension appears quite real and not merely apparent. I see no clean exit from this impasse, and I have no more to say here. I will leave this question open and move on.

Even if unable to ascertain Quine's final and stable verdict on the topic, some form of aversion to mentalism is beyond question. In the end, however, I found no grounds or reasons behind this antipathy that were strong enough to alarm me about any serious threat to the positions and assertions that I am defending in this dissertation, where the mind and psychological discourse play the essential role, and that ultimately converge on the claim that interpretation requires rationalization

5.2.1.1.2. LAST ARGUMENT AGAINST MENTALISM

In the next chapter I will accompany Quine on his path away from mentalism and the mind and - so he claims - closer into genuine explanation and understanding. In rejecting significant merits to intentional accounts, Quine is urged all the more to look for alternatives. He commits himself to contributing, within his means, to the progress of an ever more explanatory, scientific and (Quinean-)naturalistic account of linguistic facts and behavior.

It is elucidative to see how in *Mind and Verbal Dispositions*, a text from 1975, Quine distinguishes "three levels of purported explanation, three levels of depth: the mental, the behavioral, and the physiological" (Quine, 1975b, 87).

> The mental is the most superficial of these, scarcely deserving the name of explanation. The physiological is the deepest and most ambitious, and it is the place for causal explanations. The behavioural level, in between, is what we must settle for in our descriptions of language. . . . It is here, if anywhere, that we must give our account of the understanding of an expression, and our account of the equivalence that holds between an expression and its translation or paraphrase. These things need to be explained, if at all, in behavioural terms: in terms of dispositions to gross behaviour. (ibid.)

What is going on when people produce sounds and written marks and react to such productions in particular ways? Quine wants an answer to this question that does not employ intentional notions nor the usual kind of uncritical talk about *meanings*, *grasping meaning*, *sameness of meaning*. "[T]he whole point of Quine's naturalistic conception of language is precisely that our ability to speak a language can be described without using any semantical concepts at all" (Kemp, 2012, 57).

Convinced that, given the current state of scientific development, we are far from a satisfying answer of a deeper sort - physiological, neural, maybe even a physical one - Quine rests, provisionally, in the project of a behavioristic and dispositional account of the linguistic phenomenon. Linguistic beings exhibit certain dispositions to produce linguistic responses in specific circumstances and Quine wants to account for them. Within the modest limits of his speculative approach, he tries to sketch an account of their origins and causes, and of the possibility of such dispositions being shared between speakers of the same community.

These proposals, in a sense, constitute an ultimate type of argument against mentalism. To Quine's mind, any sort of development capable of bringing us closer to a (Quinean-)naturalistic account of the subject at hand - in this case, the linguistic phenomenon - would also bring us closer to proving the superfluity of a lower grade, mentalistic, explanation of the same subject. I will defend the view that Quine fails to generate any

convincing progress towards that ideal explanation of language and, accordingly, that his positive proposals pose no sort of threat or challenge to the pertinence of mentalism.

Before we come to that, however, I must return to the second family of problems concerning Quine's naturalism that was introduced above.

5.3. THE NORMATIVITY OBJECTION

There is a second important criticism of Quine's naturalism and, in particular, of his project of naturalization of epistemology that also concerns our topic and that, as such, I would like to inspect more closely. According to some critics, the re-conceived discipline now improperly neglects the *normative* aspects of belief, knowledge and epistemology.

The normative dimension in epistemology is often explained by reference to the fact that epistemology has been, and should be, thought as issuing not only *descriptive*, but also *evaluative* and/or *prescriptive* judgments. The traditional epistemologist aims not only to describe and understand the epistemic states of some thinker, and the processes through which they are brought about, but also aims at evaluation - "Are the beliefs good, justified, rational?" - and guidance - "Ought I/the thinker hold such beliefs?". Naturalized epistemology, the objection goes, loses sight of these crucial dimensions, and is confined to neutral description and factuality. It seems fit to account for any causal relations involved, but intrinsically blind to the logical or rational ones which form the natural and essential element of our fundamental concerns about knowledge and other epistemic states.

The classic reference for this objection is Jaegwon Kim's 1998 text, *What is 'Naturalized Epistemology'?*[109]. As Kim puts it, Quine "is asking us to set aside the entire framework of justification centered epistemology. That is what is new in Quine's proposals. Quine is asking us to put in its place a purely descriptive, causal-nomological science of human cognition" (Kim, 388). But, according to him, this is not how we must understand epistemology.

> We are given to understand that in contrast traditional epistemology is not a descriptive, factual inquiry. Rather, it is an attempt at a "validation" or "rational reconstruction" of science. Validation, according to Quine, proceeds via deduction, and rational reconstruction via definition.

[109] There is also Hilary Putnam's 1982 *Why Reason can't be Rationalized*. Putnam returns to the same topic in his 1988 *Representation and Reality.*

> However, their point is justificatory - that is, to rationalize our sundry knowledge claims. So Quine is asking us to set aside what is "rational" in rational reconstruction.
>
> Thus, it is normativity that Quine is asking us to repudiate. (Kim, 1988, 388-9)

Kim rejects Quine's proposal because he believes naturalization implies that "epistemology is to go out of the business of justification" (ibid., 389) and that "for epistemology to go out of the business of justification is for it to go out of business" (ibid., 391). Justification is just a paradigmatic expression of all that is arguably missing in Quine's naturalized proposal. The field of inquiry is no longer seen as a logical space and is hence stripped of its originally essential features. Notions such as reason, rationality, evidence, justification, seem to lose their place in this new program, or to reappear transformed beyond recognition - as Kim claims to be the case with evidence: "a strictly nonnormative concept of evidence is not our concept of evidence; it is something we don't understand" (ibid.). The original perspective, common to the epistemic agents' themselves, of the epistemic states and affairs, and the original internal concerns and inquiry about the rational adequacy of the agents' reasonings and states appear just as abandoned.

5.3.1. QUINE'S RESPONSE

Quine, on his part, rejects the normativity objection. He holds that a naturalized epistemology still preserves the normative aspect of the traditional discipline. He explains that his critics

> …are wrong in protesting that the normative element, so characteristic of epistemology, goes by the board. Insofar as theoretical epistemology gets naturalized into a chapter of theoretical science, so normative epistemology gets naturalized into a chapter of engineering: the technology of anticipating sensory stimulation. (Quine, 1990, 20; see also Quine, 1995a, 49-50)

According to Quine, the normative dimension in the study of knowledge is now transformed into an empirical and applied study of ways in which to improve the quality of people's knowledge and beliefs. See, for instance, here:

> Naturalization of epistemology does not jettison the normative and settle for the indiscriminate description of ongoing procedures. For me normative epistemology is a branch of engineering. It is the technology of truth-seeking, or, in a more cautiously epistemological term, prediction. Like any technology, it makes free use of whatever scientific findings may suit its purpose.

> It draws upon mathematics in computing standard deviation and probable error and in scouting the gambler's fallacy. It draws upon experimental psychology in exposing perceptual illusions, and upon cognitive psychology in scouting wishful thinking. It draws upon neurology and physics, in a general way, in discounting testimony from occult or parapsychological sources. There is no question here of ultimate value, as in morals; it is a matter of efficacy for an ulterior end, truth or prediction. The normative here, as elsewhere in engineering, becomes descriptive when the terminal parameter is expressed. (Quine, 1986b, 664– 665)

And here:

> Normative epistemology is the art or technology not only of science, in the austere sense of the word, but of rational belief generally . . . Podiatry, appendectomy, and the surgical repair of hernias are technological correctives of bad side effects of natural selection, and such also in essence is normative epistemology in its correcting and refining of our innate propensities to expectation by induction. A vest-pocket specimen of this is the exposure and correction of the gambler's fallacy: the insidious notion that a run of bad luck increases the likelihood that the next try will win. (Quine, 1995b, 50)

How good a replacement or actualization of the old normative dimension is Quine offering us here is an unsettled matter in current discussion[110]. Any definite answer to this question must be based on a firm and elaborate stance on what epistemology should be. Whether or not a Quinean epistemology is able to address and satisfy the set of traditional worries about the rationality and justification of one's beliefs is only relevant once one has decided that those are genuine and fruitful worries and concerns. Needless to say, this is not the time or place to attempt a general and detailed stance on epistemology. Likewise, I will not engage in a proper account of the complicated ongoing discussion about these issues. Even so, I want to state that I agree with those who think that there is something palpable and important that is lost in translation. Quine assures us that he still contemplates the possibility of intervening in the process of belief formation, so as to bend it towards some desired end or effect. But, if I understand him correctly, such interventions are now to be externally dictated and supported. It is no longer rationality *within us* that incites us and tells us what to believe. Truth (or, sometimes, prediction), according to Quine, is *"the ulterior end",* or *"the terminal parameter"*, and any norms or methods one might care to follow in its pursuit must be thought of as merely instrumental, not as simply valid, or rationally dictated. At the same

[110] See, for instance, Kelly, 2014, Sinclair, 2014, Rysiew, 2016, Gregory, 2008, Wrenn, 2006 and Bonjour, 2009.

time, what those norms and methods are, now that the question is some external goal and not rationality, is not something to be decided by reflection, involving the epistemic agent's intuitive apprehension of such norms and adhesion to such methods, but by the best science available, often beyond the reach of that agent.

Not aiming at a complete treatment of this discussion, nor at a conclusive verdict on the merits and problems of naturalization of epistemology, I take, instead, a much more modest task and goal. In the next section I will try simply to reveal how interpretation involves epistemology of a sort that does not fit Quine's model, even considering his latest protests and suggestions. I will also defend the view that this is a tenacious sort of epistemology, one that we could not easily renounce.

5.3.2. NORMATIVE EPISTEMOLOGY IN INTERPRETATION

There is epistemology involved in language acquisition and interpretation. Davidson himself is quite explicit about it in declaring that "the methodology of interpretation ... is nothing but epistemology seen in the mirror of meaning." (Davidson, 1975, 169). As we saw, the method of the radical interpreter is that of getting meaning ascriptions via belief ascriptions. However, since he also lacks any direct access to the relevant beliefs he has no other option but to project himself into the speaker's perspective and search for what would be rational to believe *from there* - both in the way in which those beliefs *must* agree with what is going on in the world, and in the way in which they *must* relate to one another and to other propositional attitudes so as to constitute a solidary system. The interpreter's projection into the other person's shoes can serve his purpose only if both interpreter and speaker find themselves under the same general epistemic obligations, under the same rationality constraints - here is where we find the mirrored epistemology. We can say that where the epistemologist starts with the beliefs and the relevant evidence so as to evaluate the rationality of the process leading from the latter to the former, the radical interpreter must start only with the evidence, presuppose the rationality of the process and, that way, discover the beliefs.

The epistemology involved in interpretation, however, is of the traditional brand, not of the naturalized sort. Of course, I do not mean to suggest that Davidson has attempted in any way to resume either the *conceptual* or the *doctrinal* projects described above.

Traditional epistemology is not exhausted in the two-pronged project described by Quine[111]. What I mean is that the notions, worries, questions and interests are closer to those of non-naturalized epistemology.

The notion of *rational reconstruction* is exemplary of the traditional, normative approach. Quine uses it both in a strict sense - to talk about the Carnapian program[112] of "construction of physicalistic discourse in terms of sense experience" (Quine, 1969a, 75) - and in a less stricter sense, so as also to fit other endeavors that are analogous in the relevant sense - as when he says, for instance, that "[i]n the old epistemological context the conscious form had priority, for we were out to justify our knowledge of the external world by rational reconstruction, and that demands awareness" (Quine, 1969a, 84). Personally, I think we have good reasons to understand the interpreter's task as *analogous in that relevant sense.*[113] What interests Davidson and his interpreter is the rational aspect of the processes of knowledge and belief formation. From this perspective, the only environment fit to accommodate and account for a belief is a mental system made out of rational connections - evidential, conceptual and logical connections. This is the system that we expect the interpreter to able to (partially) reconstruct and elucidate, and so I sustain that *rational reconstruction* is a title suited for it.

The sort of engineering of epistemic states that Quine points to in his response to the *normativity objection* is a very different exercise from that of critically evaluating someone's

[111] This confusion has prompted a different type of objection to Quine's program of naturalized epistemology. Rysiew (2016) calls it "the non-sequitur objection", denouncing Quine's improper identification of traditional epistemology with radical foundational projects such as Cartesian epistemology or Carnap's reductionism. Consequently, the objection goes, "whatever the merits of Quine's attack on the sort of strong foundationalist program practiced by Descartes and the local empiricists, they fail to motivate any rejection of [traditional epistemology] as such" (ibid.).

[112] The classical reference is Carnap, 1928, *The Logical Structure of the World*. There we find the first extended development and application of the notion. For a broader understanding of the complex role the notion plays in Carnap's work see Friedman, 2007, Uebel, 2007 and Demopolous, 2007.

[113] Quine claims the expression for yet a third use - for instance, here: "It is rational reconstruction of the individual's and/or the race's actual acquisition of a responsible theory of the external world. It would address the question how we, physical denizens of the physical world, can have projected our scientific theory of that whole world from our meager contacts with it: from the mere impacts of rays and particles on our surfaces and a few odds and ends such as the strain of walking uphill.

Such is my option. It is part and parcel of empirical science itself, with rational reconstruction intruding only at the conjectural interstices or where complexities of historical accident becloud the schematic understanding we are seeking." (Quine, 1995a, 16)

I do not find this third use very fitting. The rationality involved is outside the processes. Here we find no consciousness or awareness - nor even the possibility of such consciousness or awareness - of the processes that result in knowledge, let alone a sense of their rationality. Or, as Davidson puts it, "[w]hen we try to understand the world as physicists, we necessarily employ our own norms, but we do not aim to discover rationality in the phenomena" (Davidson, 1991, 215).

epistemic path as it is, or can be, presented and realized in the first-person. By this I mean the reflection about the rational quality or adequacy of the grounds and reasonings supporting someone's beliefs; the attempt to represent someone's condition and conduct as a thinker and a believer, so as to make sense of that condition and conduct. This, I claim, is fairly described as an epistemological effort, of a very common and pressing type, and with a special emphasis on evaluation and prescription.

It is undeniable that there is such a perspective of mental states and operations. On a regular basis, humans take this perspective both with relation to their own beliefs but also with relation to the beliefs of others. We often reflect upon our own beliefs and states and adjust them, or not, in conformity. We do the same with regard to the epistemic states of others. We consider whether someone is justified in believing something, whether someone perceives something from the place where she is standing, whether someone is in possession of all the relevant information or is likely to be misled by planted evidence, and so on.

Irrespective of whether we should voluntarily continue developing this type of inquiry as a scientific or philosophical discipline or not, it appears that we cannot help engaging in it in the course of our natural and intuitive interactions with other humans, and reflection upon our own thought. *Natural* or *folk* epistemologists do not have to be perfectly explicit and clear about the epistemological nature of their exercise, and they are (usually) not aiming to convince a skeptic. Still there is epistemology involved, there is a flagrant appeal to intuitive epistemic standards in such episodes, the rationality and likelihood of certain beliefs is measured, and the results inform the inquirer's own states or ascriptions.

In a very recent volume, David Moshman (2014) reviews the literature from a number of diverse areas, or enterprises, in cognitive psychology to reveal how each contributes to the identification and understanding of various aspects of humans' natural epistemological thinking and practices. The common ground for all these approaches, before any eventual disputes and polemics, and the starting point for further elaboration, is the fact of there being a pre-theoretical and intuitive concern with the quality of one's beliefs at work in everyone's everyday cognitive life where normative questions and notions such as rationality and justification take the leading role. This book provides a very convenient and up-to-date map and introduction to a vast and scattered body of work and theory that investigate and document these facts and issues.

Even without getting into the sophisticated and theoretical proposals such as the ones

explored in the volume mentioned - such an incursion is beyond the scope of this dissertation - common sense should be more than enough to allow us to recognize the assiduous presence of considerations of justification and rationality determining or regulating thoughts and thought ascriptions. Think again, for instance, of the puzzle and of the puzzle-solving perspective presented in the introduction. There is obviously an internal angle to the problem, one where the child's epistemic perspective is considered and a solution is arrived at by having him subject to certain constraints of rational and justified belief, i.e. by applying an evaluative and prescriptive sort of epistemology to his mental states. This is so, and will not cease to be so, no matter how many interesting, alternative and external angles - be they biological, neural, physical, or whatever - may also be found and developed over the same problem.

This internal angle is analogous to the one the radical interpreter must assume, as detailed a few paragraphs above. This internal angle is also perfectly patent in the empirical literature and experimental studies on word learning presented in Chapter 4. Whatever the more specific processes or methodologies involved, the mindreading efforts required for the task are ultimately regulated by the need to reveal the general and common rationality of speakers. Children have no alternative but to resort to their own nascent epistemic standards to infer the beliefs, and other states, of their interlocutors. I should also repeat that the employment of mindreading and rationalizing skills is, of course, not confined to the task of word learning but, instead, seems present, from very early on, in most forms of human interaction.

Rational reconstruction, I defend, is not competing to elucidate the same facts or the same aspects of objects and events as natural sciences[114], and I see no reason for not adopting an ecumenical stance here, accepting the pertinence of different conceptions of epistemology, of referring to different conceptions of justification, and of describing different aspects of the processes of belief formation[115].

Quine, at times, seems to think of naturalized epistemology as being the only worthy way of doing epistemology, of inquiring and assessing the epistemic states of a thinker. I

[114] I take Davidson's side: "Whether the features of a psychological theory I have been rehearsing … show that a psychological theory is so different from a theory in the natural sciences as not to deserve to be called a science I do not know, nor much care. What I am sure of is that such a theory, though it may be as genuine a theory as any, is not in competition with any natural science." (Davidson, 1995a, 134)

[115] See Bonjour, 2010.

oppose such a view and defend the interest and likely indispensability of the traditional way of going about it, in particular, inasmuch as this type of epistemology is required for language learning and interpretation.

5.4. QUINE'S *COMMON INTERPRETER*

In the same spirit with which we recognized a certain complexity and ambivalence in Quine's appreciation of the explanatory potential of the mind and of the mentalistic idiom, we must also not neglect the fact that Quine's attention to language is not entirely limited to the natural aspect of the phenomenon. Besides physical bodies, organisms, brains, nervous systems, dispositions, sensory inputs and other prototypical figures of Quinean naturalism, there are other presences populating his views and texts on language. We can synthesize these, I propose, in the idea of *the common interpreter.* The common interpreter is a rational agent and thinker, endowed with a perfectly ordinary mind and worldview - that is, typically human in the broadest of outlines - of a kind which he, in his turn, intuitively presupposes in his fellow humans and interlocutors. This common interpreter is practically indiscernible from Davidson's interpreter in the matters that we are occupied with: he must engage in folk psychology and epistemological inquiry of the traditional brand in order to acquire a language, to interpret and to communicate.

Despite his low profile, the common interpreter is never completely forgotten throughout the whole of Quine's work. Quine did not begin to acknowledge him, and to appeal to him, only in his later texts when - as we will see - he withdraws to a more realistic conception of what is essential to translation. No, he is there from very early on. He is there, significantly, in 1960, in the first paragraph of *Word and Object*.

> Each of us learns his language from other people, through the observable mouthing of words under conspicuously intersubjective circumstances. Linguistically, and hence conceptually, the things in sharpest focus are the things that are public enough to be talked of publicly, common and conspicuous enough to be talked of often, and near enough to sense to be quickly identified and learned by name; it is to these that words apply first and foremost. (Quine, 1960, 1)

However, we also find him in Quine's last book, *From Stimulus to Science*, from 1995,

> The linguist will rely also on observation of the local folkways. The child does too, but the

> linguist is a more seasoned observer. Unlike the child, the linguist will not accept everything the native says as true. He will indeed assume sincerity, barring evidence to the contrary, but he will try as an amateur psychologist to fit his interpretations of the native's sentences to the native's likely beliefs rather than to the facts of circumambient nature. Usually the outcome will be the same, since people are so much alike; but his observation of the folkways is his faltering guide to the divergences. (Quine, 1995a, 80)[116]

The appeal to this interpreter implies, I believe, the implicit and inadvertent recognition of pertinence for the kind of epistemology and vocabulary that Quine repudiated. Furthermore, if such a figure, the common interpreter, comes to reveal itself as something more than a mere heuristic device, if it comes to reveal itself as indispensable to language understanding and to any fair description of its acquisition process, so will the kind of epistemology and idioms he resorts to.

5.5. CONCLUSION

I am defending here that interpretation requires rationalization. Rationalization makes an essential use of mentalistic idioms and explanations, and it incorporates an epistemological investigation of the non-naturalized sort. After examining the two Quinean challenges - Quine's anti-mentalism, and his promotion of the naturalization of epistemology - I conclude that there is no real, strong case either against the methodology that I defend is required for interpretation or, in any other way, against the claim that this methodology is required.

Regarding the first challenge, Quine's objections to mentalism, first of all it must be noted they are not directed at its quotidian use, and do not dispute the practical value of that type of discourse and explanation. Accordingly, Quine does not actually question the need or adequacy of rationalization in *ordinary* interpretation.

Second, there is the question of Quine's harsh rejection of scientific value to disciplines or enterprises that, in one way or another, refer to mental realities or employ

[116] See also: "Early and late I recognized empathy as the strategy in radical translation. My use of the word "empathy" is only recent and has been noticed, but I had already recognized the radical translator's approach as empathetical in *Word and Object* and indeed nine years before. "The lexicographer," I wrote, "... depend[s] ... on a projection of himself, with his Indo-European Weltanschauung, into the sandals of his Kalaba informant." It is by empathy that we estimate our interlocutor's perceptions. Their neural implementation is as may be." (Quine, 2000c, 410)

mentalistic explanations - where one finds included, in particular, the sort of account of word learning, congenial to Davidson's take on interpretation, that was presented in Chapter 4. I oppose Quine's rejection and, in response, note that the arguments produced appear too feeble to seriously threaten well-established and fruitful disciplines, fields of inquiry or methodologies.

Third, and finally, there is the crucial remaining disagreement between Quine's position and what I am defending here. It resides in Quine's expectation of some form of ideal and general overcoming of all folk psychological methods by (Quinean-)naturalistic disciplines and investigation, whether or not that progress should come to be reflected in any change in humans' everyday conduct. Such an expectation, as noted above, creates a tension, first of all, internal to Quine's own system. Quine's apparent endorsement of the anomalism of the mental is hard, if not impossible, to reconcile with his ideal of the continuity of science and of naturalistic hegemony. But, most importantly, it also challenges, however remotely and merely programmatically, the thesis defended here that there is no way to interpret outside rationalization. Quine's main argument - if we may call it that - in favor of the possibility of such an overcoming of folk psychology by natural sciences takes the form of an actual first sketch of a (Quinean-)naturalistic account of language and linguistic practices, with an appendix on naturalistic methods of translation or interpretation. The ground was prepared in this chapter, but Quine's positive proposals will only be considered in the next. Any success by Quine in bringing us closer to a viable naturalistic account of language and, especially, of a non-rationalizing method of interpretation would, to some degree, compromise or refute our thesis. However, I will defend that he does not achieve any relevant success.

The second challenge is generated by Quine's extension of his naturalism to epistemology. Quine advocates a radical revision of the discipline. In response, and following a by-now classic line of criticism, I note that the new discipline loses sight of crucial aspects of the traditional epistemological inquiry.

In addition to this, I highlight the existence of a non-specialized and non-theoretical face of epistemological thinking and practices, and I point to its essential role in rationalization and interpretation. This natural epistemology seems safe from furors of naturalistic revision. Ordinary epistemic agents and interpreters are naturally taken by epistemological concerns over the epistemic quality of their own and others' epistemic states,

and they have no other way to address such concerns except through a normative inquiry that explores their more or less spontaneous and implicit conceptions of rationality and justification.

6. A Naturalistic Alternative to Rationalization

The first part of this chapter is dedicated to a critical presentation of Quine's proposals for a (Quinean-)naturalistic understanding of language and linguistic practices. I confine the reflection to a specific and pivotal class of sentences - observation sentences. Quine maintains that speakers are conditioned to respond to these sentences in specific ways given certain stimulatory conditions. He looks for behavioral criteria suited to the identification of observation sentences. He postulates a certain similarity relation, *perceptual similarity*, and affirms it to hold among stimulations that prompt like responses to observation sentences.

I maintain that Quine's rudimentary account is inadequate. I accept Quine's criteria for the identification of observation sentences but assert that no sentence satisfies it. This is my central point against Quine's general and sketchy proposal. Without observation sentences, his whole explanation collapses. I also state some doubts about the naturalistic credentials of perceptual similarity, and stress that Quine provides no further elucidation of the notion nor of how it gets instantiated.

In the second, and shorter, part, I consider several potential Quinean methods of translation and interpretation in search of a naturalized alternative to interpretation via rationalization. I find no credible alternative that genuinely dispenses rationalization.

6.1. NATURALISTIC ACCOUNT OF OBSERVATION SENTENCES

Let us now try to understand what kind of naturalistic elucidation Quine hopes to obtain for the linguistic phenomenon. Concerning a certain type of elementary linguistic behaviors, what Quine has to offer is an account of such behaviors as responses to sensory stimulations. We, language users, have been inculcated with certain dispositions to react with certain linguistic responses exclusively on the basis of certain patterns of activation of our nerve endings[117]. Those are the dispositions that manifest themselves when we produce or

[117] I will be using 'sensory stimulations', 'triggering of sensory' receptors, 'activation of nerve endings',

react to those expressions that Quine refers to as *'observation sentences'*.

Observation sentences are also *occasion sentences* - sentences whose truth or falsity varies according to the circumstances - plus a certain specifying difference. But now compare

a. Look, there's your neighbors' dog.

with

b. Look, there's a dog.

Both are occasion sentences, but the second is better suited to double also as an observation sentence. The intuitive idea behind it is pretty obvious: observation sentences are such that every speaker has to be able to tell whether that sentence is true at that time and place just from what he can see going on around him. What matters now is whether we can really make sense of this idea in naturalistic terms, that is, if we can naturalistically discern and understand this special class of sentence.

Their central feature, and the reason Quine pays them particular attention, is supposed to be their immediate link to stimulation. Those are the "sentences that are directly and firmly associated with our stimulations" (Quine, 1990, 3), the sentences whose "distinctive trait is the sufficiency of present impingements" (Quine, 1974, 40).

> Observation sentences - as I call them - can be conditioned outright to distinctive ranges of sensory intake, or as physicalists let us say *neural* intake. The child can be conditioned simply to assert or assent to the sentence under some distinctive stimulation… (Quine, 1993, 108)

According to Quine, these are the sentences by which speakers must start language acquisition and translation. These are also the sentences by which we start the exploration of Quine's naturalistic account of language. If, in fact, we can find in Quine any serious candidate to something akin to an effective elucidation, delivered in naturalistic terms, about the workings and nature of language, it will have to be about observation sentences. Those are the sentences best placed and best suited by the rudimentary kind of explanation he has to offer.

With the notions of *stimulation* and *response to stimulation* playing the central role in the explanation, verbal responses involving observation sentences become the paradigmatic

---

'sensory input' and 'neural input' interchangeably. In doing that, I believe I will not be missing any distinction important to the matters in hand.

objects of the account since this is where the connection between stimulation and response is the most direct. The bigger the gap between stimulation and response, the more appealing becomes the employment of mental states and reasoning in their mediation, and the more fragile the plausibility of a fully naturalistic account of the episode.

Our crucial question is the contrast between the understanding of linguistic behavior as immediate, brute, causal reactions to stimulations, and the understanding of those same behaviors as rational and intentional answers. Certainly, a full and satisfactory account of language in accordance with Quine's naturalism would have to be able to reveal all linguistic behaviors - not only those involving observation sentences - as fully naturalistically accountable events. That is, the dispensability of mind and reason in such an account must be true for all language. However, these being the first and the easiest sentences to explain, it is here that Quine's account is expected to make the superfluity of such notions most sensible. On the other hand, if, for some reason, we find that the account does not work even at this level, we can assume that it does not work at all.

For observation sentences*,* the correlations between exposure to certain stimulatory patterns and the coming about of certain linguistic responses should be direct. In particular, mental states must not be appealed to in determining and supporting these correlations. If Quine is right, we do not have to think of such behaviors as actions determined by communicative intentions - as a result, say, of some desire to express a certain belief taken together with the belief that the utterance of the sentence in question, given what one believes to be its meaning, is an adequate means for that. Alternatively, we can think of, and explain, those behaviors simply as natural events, and as the natural result of previous natural events.

The existence of observation sentences is an indispensable piece of Quine's understanding of language. Those are the sentences through which language and theory are linked to the world. Non-observational sentences only have a content and a use through their links to the observational ones. They secure both the learnability of the language and the agreement necessary for every sharing or parting of opinions or theories. These are the sentences that, as unstructured wholes, can be learned - and some necessarily are - by direct conditioning to certain stimulatory conditions. Simplifying it a little, we can say that every such sentence is repeated to the child or language learner by some community member, usually the parents, when that person detects that the child's stimulatory situation is appropriate to the acceptance of that sentence. These are the sentences that constitute the

basis of any understanding, communication and discussion. For instance, they "are just the sentences on which a scientist will tend to fall back when pressed by doubting colleagues" (Quine, 1960, 39). In a sense, they report the most primitive and simple empirical facts and as such they must generate "firm agreement on the part of well-placed observers" (ibid.). However, let me insist, observation sentences are only apt to play these roles because of their allegedly being *directly* and *firmly* keyed to stimulations. The viability of Quine's naturalistic conception of language is dependent upon the existence of such strong and direct links. Quine gives us no way to conceive of more complicated relations between language and world, between verbal behavior and stimulations, that does not appeal to these simple relations between verbal behavior and stimulations.

In sum, then, this is Quine's naturalistic response to the general foundational question in semantics - *What makes it the case that some expression means what it means?* - or, in more neutral terms, aiming to comply with Quine's rejection of folk psychology and semantics - *What makes it the case that some expression has the semantic value it has*? According to Quine, the semantic value of an observation sentence is determined by the anchoring of assent and dissent to that sentence to particular stimulatory conditions. The semantic value of each remaining expression is a much vaguer business but it is determined, somehow, by the inferential relations between that expression and some high number of other expressions in the language, including, in particular, observation sentences. In what follows we will explore more closely the details of his proposal for only the first and pivotal class of sentences.

6.1.1. PERCEPTUAL SIMILARITY

Quine proposes a certain relation, *perceptual similarity*, to help us make sense of these sentences and of their links to stimulations. Stimulations prompting the linguistic behavior of acceptance or rejection of an observation sentence must have something in common. To be a competent user of that sentence requires the ability to detect this commonality between stimulatory episodes. These episodes must be perceptually similar.

This similarity is to hold between *global stimuli.*

> What must be compared are episodes of global stimulation: the set of all the sensory receptors triggered at a given moment, or the temporally ordered set of all the sensory receptors triggered during

> some relevant lapse of time. (Quine, 1995a, 17)

Global stimuli can be *receptually similar*.

> Each global stimulus is an ordered subset of the subject's nerve endings, and two such subsets are more or less similar in the obvious sense according as they comprise more or less the same nerve endings in more or less the same order. This I call receptual similarity. (ibid.)

Or, more simply:

> Episodes are receptually similar to the degree that the total set of sensory receptors that are triggered on the one occasion approximates the set triggered on the other occasion. (Quine, 1974, 16)

This sounds terribly crude, with its oversimplified understanding of nerve endings' reactions to inputs, and its sketchy and doubtful strategy for comparing global sets of such triggerings. Moreover, receptual similarity cannot, in any case, be the kind of similarity that interests us. The same observation sentence can be uttered properly in different circumstances where the global stimulus diverges greatly, receptually speaking. We need no more than the merest hint to make this point obvious enough: compare a cat seen at night in an alley and a cat lying on the grass at noon.

Perceptual similarity contrasts with receptual similarity, and is meant to articulate the possibility of finding global stimuli alike with respect only to some privileged portion of the input, and only *from a certain angle* that the perceiver must somehow be sensible to. It is not because we *receive* but because we *perceive* something similar on different occasions that we are disposed, on those occasions, to accept the same sentences. It is not the whole situation that has to be similar, just some particular aspect of it.

Whatever that perceptual similarity turns out to be, the relation is linked - I would say *by definition* - to observation sentences: "perceptual similarity relates the episodes that warrant assent to an observation sentence." (Quine, 1974, 43) Grasping the first is the condition for mastering the second.

> The learning of an observation sentence amounts to determining, as we may say, its similarity basis. By this I mean the distinctive trait shared by the episodes appropriate to that observation sentence; the shared trait in which their perceptual similarity consists. (ibid.)

We must beware of understanding perceptual similarity as a subjective notion grouping together stimulations that the subject somehow experiences as similar[118]. Quine actually refers to perceptual similarity as the *'physical analogue'* of a Carnapian phenomenological notion of *experience part similarity* (Quine, 1995a, 17). Any appeal to subjectivity would fail to serve Quine's ambition to explain language without employing the mind. Quine has to hold on to the idea that the similarity in question must also lie *out there*, in the physical triggering of nerve endings. The property or relation is meant as a physical or (Quinean-)naturalistically describable one. I take this to mean, at least, that the similarity in question must be directly *readable* from some external and objective description of the patterns of stimulation.

We now have a posited relation and a name for it but, clearly, still no progress in any natural explanation of the uses of observation sentences nor of the relation holding together - that is the hypothesis - the stimulations that prompt responses to each such sentence. What more has Quine to offer?

Quine distinguishes two steps, corresponding to two levels of success, in the task of providing a naturalistic account of observation sentences and perceptual similarity. The more modest one is that of providing an objective test capable of proving the objective presence of this similarity relation - and hence of observation sentences - on specific occasions. The more ambitious one is that of actually explaining the relation, that is, of elucidating the facts and mechanisms behind its constitution and detection by an individual.

Quine clearly attempts the first one, and seems convinced he has worked it out satisfactorily. I will consider this first part of the job in the next sections and explain why I think Quine fails at it. I leave the second part for after that.

---

[118] Here, for instance, we find a similar alert. "The color of the object is salient in the one episode because of its brilliance and saturation. and the shape is salient in the other episode because of boundary contrast or movement. It would be intolerable to deprive ourselves of these quick and vivid ways of speaking. But let us remember that this is all meant to be, in the end, a matter of physiological mechanisms, manifested in behavior." (Quine, 1974, 25-6)

6.1.2. SIGNALING OBSERVATION SENTENCES AND PERCEPTUAL SIMILARITY

I will deal here with perceptual similarity only inasmuch as it is supposed to be the relevant relation between stimulations that prompt responses to observation sentences. Consequently, I will eventually neglect certain aspects and instances of the relation[119], and focus only on the identification of perceptual similarities that account for responses involving observation sentences.

Quine, throughout the years, kept trying to sharpen a good behavioristic criterion for the detection of perceptual similarity and observation sentences. He tried several distinct approaches but, as I will defend, he never quite accomplished what he was looking for. As I see it, this is not so much because the tests are not good, but because there seem to be no observation sentences in the Quinean sense. That is, there seem to be no sentences that speakers will respond to based only on how their sensory receptors are stimulated.

Given that stimulation is permanent, every piece of verbal behavior counts, at first, like a candidate observation sentence. How, according to Quine's proposal, is one to tell, behavioristically speaking, whether that piece of verbal behavior stands or not in a direct and firm relation to the sensory stimulations the speaker has undergone a moment earlier, or whether a more complicated relation between it and that speaker's stimulatory career is the case?

I will run through the most significant statements of Quine's proposed tests, considering their points and merits, and highlighting continuities and differences between them. There is a first distinction between such tests that Quine makes explicit, the distinction between public or social criteria and private or individualistic ones. In each of these categories I mark a further distinction - a different and unrelated distinction in each case.

6.1.2.1. SOCIAL CRITERION A

Quine's original proposal, presented in *Word and Object* (1960), belongs in the public criteria category.

---

[119] There are other, actual or possible, perceptual similarities to which no observation sentence corresponds but the opposite cannot be true.

> Occasion sentences whose stimulus meanings vary none under the influence of collateral information may naturally be called *observation sentences* ...
>
> ...in behavioral terms, an occasion sentence may be said to be the more observational the more nearly its stimulus meanings for different speakers tend to coincide. (Quine, 1960, 43)

The method suggested for the identification of observation sentences employs another technical notion, that of *stimulus meaning* (Quine, 1960, 29). The stimulus meaning of an observation sentence is composed of two sets: its affirmative stimulus meaning - simplifying it somewhat, the set of sensory stimulations that elicits assent to that observation sentence - and its negative stimulus meaning - the set of sensory stimulations that elicits dissent to that observation sentence.

The test asks us to compare the stimulus meaning of occasion sentences and to count a sentence as observational when its stimulus meanings are identical, or nearly so, for every speaker.

A first problem is that we lack any viable way of identifying the stimulus meaning of a sentence for a speaker. A better understanding of perceptual similarity and the ability to grasp the specific similarity relating stimulatory episodes that cause assent to a certain observation sentence would, in principle, allow us to determine its positive stimulus meaning. We do not have that. In fact, that is what we are looking for. Outside an appeal to perceptual similarity, nothing better suggests itself than the individual test of specific sensory stimulations so as to separate the range that causes assent and the range that causes dissent for some particular sentence. Of course, the exhaustive test of every possible stimulation is a remote chimera but, and although Quine gives us no idea about the terms in which to describe or refer to sets of sensory stimulation, there is some residual plausibility to the idea of being able to somehow collect some of the sensory stimulations that belong to the stimulus meaning of a certain sentence for a certain speaker. Then one would still have to compare them with stimulations generating like responses in other speakers to see if they match. It is an unpromising strategy but, as I have said, there is at least some residual plausibility to it[120].

---

[120] Quine is aware of the practical impediments to such a test. "In taking the visual stimulations as irradiation patterns we invest them with a fineness of detail beyond anything that our linguist can be called upon to check for. But this is all right. He can reasonably conjecture that the native would be prompted to assent to 'Gavagai' by the microscopically same irradiations that would prompt him, the linguist, to assent to 'Rabbit', even though this conjecture rests wholly on samples where the irradiations concerned can at best be hazarded merely to be pretty much alike." (Quine, 1960, 28)

Lacking an operative notion of perceptual similarity the only way of (tentatively) approaching sameness of stimulus meaning is via sameness of stimulations. Accordingly we can reformulate the test in terms of stimulations. We can say instead, to the same effect, that *occasion sentences to which distinct speakers always give the same response when sensorily stimulated in the same way are observation sentences.*

How does it work? i.e. how is it supposed to be that general agreement in verbal responses to stimulations should signal the fact that such responses are firmly and directly correlated to those stimulations? Here is what I take to be the reasoning behind this test: starting with the following three assumptions:

- Speakers share their *knowledge of the language* - i.e. we assume that "language as a socially inculcated set of dispositions is substantially uniform over the community" (Quine, 1960, 40) - but that all other beliefs and knowledge can vary at will.
- If a sentence is an observational one, verbal responses involving it are the result of the speaker's knowledge of the language and of the prompting stimulation.
- If a sentence is a non-observational one, verbal responses involving it are the result of the speaker's knowledge of the language, of the prompting stimulation, plus whatever particular beliefs and collateral information on the topic each speaker may possess and not share with all others.

One must conclude that:

- Since the variable factors affect only responses to non-observational sentences, general agreement must indeed signal firm and direct correlations.

6.1.2.2. SOCIAL CRITERION B

This social criterion evolved in later texts as Quine became uneasy about the requirement of shared stimulation.

> ... an event of stimulation … is the activation of some subset of the subject's sensory receptors. Since the linguist and his informant share no receptors, how can they share a stimulation? We might say rather that they undergo similar stimulation, but that would assume still an approximate homology of nerve endings from one individual to another. Surely such anatomical minutiae ought not to matter here. (Quine, 1990, 40)

Consequently, Quine stopped invoking sameness of stimulation in formulating his social test, and started exploring instead the notion of *witness to a situation*, as in the following passages:

> The requirement of intersubjective agreement already affords us just the definition we need. A sentence is observational insofar as its truth value, on any occasion, would be agreed to by just about any member of the speech community witnessing the occasion. (Quine, 1974, 39)

> ...unlike a report of a feeling, the sentence must command the same verdict from all linguistically competent witnesses of the occasion.
>
> I call them observation sentences. (Quine, 1990, 3)

> I retain my 1981 definition of observation sentence for the single speaker [121], and then account a sentence observational for a group if it is observational for each member and if each would agree in assenting to it, or dissenting, on witnessing the occasion of utterance. We judge what counts as witnessing the occasion, as in the translation case, by projecting ourselves into the witness's position. (Quine, 1990, 43)

> What I call observation sentences … They are occasion sentences—true on some occasions, false on others. Furthermore they report intersubjectively observable situations, observable outright. That is to say, all members of the language community are disposed to agree on the truth or falsity of such a sentence on the spot, if they have normal perception and are witnesses to the occasion. (Quine, 1995a, 22)

> ...the public requirement on observation sentences, namely unhesitating concurrence by all qualified witnesses. (Quine, 1995a, 44)

Some of these statements came even before Quine finally worked out, to his satisfaction, a solution to the problem of explaining otherwise the coordination among speakers. This happened in 1994, in his *From Stimulus to Science*, where he first appealed to a *preestablished harmony* allowing speakers to agree in their responses to situations independently of there being any homology between sensory receptors or identity between stimulations, thus rendering the *'anatomical minutiae'* definitely irrelevant.

> Global stimuli are private: each is a temporally ordered set of someone individual's receptors. Their perceptual similarity, in part innate and in part molded by experience, is private as well. Whence

[121] This 1981 definition is presented in the first quotation in the next section.

then this coordination of behavior across the tribe? It requires that if two individuals jointly witness one scene, and subsequently jointly witness another scene, and the one witness's global stimulations on the two occasions qualify for him as perceptually similar, usually the other witness's stimulations will so qualify for the other witness.

> So we see a preestablished harmony of perceptual similarity standards. If two scenes trigger perceptually similar global stimuli in one witness, they are apt to do likewise in another. (Quine, 1995a, 20-1)

The reasoning behind the test remains the same as with social criterion A. At most we need a broader notion of uniformity - say, appealing not to *identical* but to *equivalent* sets of stimulations - when conceiving the shared "language as a socially inculcated set of dispositions ... substantially uniform over the community" (Quine, 1960, 40)

Those are the social tests. Then there are the individual ones.

6.1.2.3. INDIVIDUAL CRITERION A

A clear formulation of what seems to me the most straightforward individual test for observation sentences can be found in "Empirical Content" (1981).

> If querying the sentence elicits assent from the given speaker on one occasion, it will elicit assent likewise on any other occasion when the same total set of receptors is triggered; and similarly for dissent. (Quine, 1981a, 25)

The reasoning behind it is simple. It explores the supposed fact that, if a strong correlation is the case between stimulations and responses, a right stimulation must be *sufficient* to generate the right response. Being a sufficient condition, the stimulation must prompt the right response every time. If a stimulation that generates a positive or negative response to a certain sentence on a certain occasion were to fail to prompt a like response to the same sentence on a different occasion, the conclusion would have to be that the stimulation was not, after all, sufficient for any of the responses - and, hence, that the link between stimulations and responses for that sentence was not a firm and direct correlation.

It is not perfectly clear, but I think we find the same idea behind this passage:

> … sentences that are directly and firmly associated with our stimulations. Each should be associated affirmatively with some range of one's stimulations and negatively with some range. The sentence should command the subject's assent or dissent outright, on the occasion of a stimulation in

> the appropriate range, without further investigation and independently of what he may have been engaged in at the time.
>
> … I call them observation sentences. (Quine, 1990, 3)

6.1.2.4. INDIVIDUAL CRITERION B

There is yet a different test that Quine frequently refers to, also belonging to the *private* category. In one of its first appearances, in *The Roots of Reference,* the test is offered as a "behavioral condition for perceptual similarity",

> ... we can correct our formulation of the behavioral condition for perceptual similarity, to read thus: *a* is shown to be perceptually more similar to *b* than to *c* when the subject has been conditioned to respond in some fashion to all episodes in the receptual neighborhood of *b,* and to withhold that response from all those in the receptual neighborhood of *c,* and is then found to so respond to those in the neighborhood of *a*. (Quine, 1974, 17-8)

A test for perceptual similarity, as we have seen, should naturally be able to be adapted to double as *a behavioral condition for observation sentences.* It might be argued that Quine attempts precisely that. For instance, when, in *From Stimulus to Science*, he states that "[t]he private requirement [to be an observation sentence] both early and late is just that the sentence be keyed directly to a range of perceptually fairly similar global stimuli" (Quine, 1974, 43), what is suggested is that the test he has provided for perceptual similarity twenty pages before, identical to the one just quoted[122], can be of some use in the identification of observation sentences. However, I do not see how. I fail to see how this test for perceptual similarity could help us determine whether or not a sentence is observational.

I think one can begin to appreciate the way in which this test is not suited for that job by noticing that it already presupposes that the responses to the stimulations are observational ones. It presupposes that the responses will be observational, and then goes on to test whether or not the third stimulation belongs to the range of positive stimulus.

---

[122] An individual's standards of perceptual similarity, at any given stage of his development, are in principle objectively testable, as follows. The individual happens to make some move on the occasion of some global stimulation, and we reward the move. Later we stimulate him again in a receptually somewhat similar way, and in view of the past reward he makes the same move again, but this time we penalize it. Finally we stimulate him again, this time in a way intermediate between the two stimulations, in respect of receptive similarity. If he makes the move a third time, despite the recent penalty, we conclude that the third stimulation was perceptually more similar to the first than to the second. (Quine, 1974, 17-8)

Supporting my point, see how misguided seems the application of the test to a non-observational property - being a bachelor, for instance - supposedly so as to exclude it: *take some speaker, reward her responses to a picture of a certain bachelor, penalize identical responses when the picture shown is of a married men, and test for perceptual similarity with a third picture.* Clearly, there is no sense in such a test and no significant result can come from it.

In talking about a "behavioral condition for perceptual similarity", there is some ambiguity - and probably even some confusion – between:

a) a behavioral condition for *enough* perceptual similarity between two stimulations along a certain dimension of perceptual similarity, and

b) a behavioral condition *for there being at all* any perceptual similarity, i.e. a test that could determine whether or not there even is a dimension of perceptual similarity along which the two stimulations are to be positioned and tested for closeness.

The test Quine proposes here concerns only condition *a*. Something along its lines should be able to tell us whether one perceives an apricot as sharing its color with lemons or with oranges. But it is only condition *b.* that is directly relevant to us when we are trying to find a way to recognize verbal responses firmly and directly keyed to stimulations.

This ends our tour of Quine's tests for observation sentences and perceptual similarity. In conclusion, I propose that we keep Social Criterion B, being simply a more liberal version of Social Criterion A, and Individual Criterion A, dropping Individual Criterion B on the grounds put forward above.

6.1.2.5. NO SENTENCE SATISFIES THE CRITERIA

Here is the amazing thing: no sentence satisfies the criteria offered for the detection of firm and direct correlation of verbal responses to the sentence and occurring stimulations. As I have said above, I do not think the problem lies with the criteria. The two tests just endorsed are adequate enough for their purpose - at least theoretically so, as they may be too impractical for actual implementation. The problem is rather that there really seems to be no Quinean observation sentences.

That is a serious problem as such sentences are both essential to Quine's picture of language and the most amenable to the type of naturalistic explanation he has sketched. Without them, it is hard to see what could be saved from Quine's theory of language and meaning - much the same for his epistemology. Without them, Quine's exhortation for a naturalistic study of language would stand even emptier of any positive suggestions about how to do it.

To see that there are no observation sentences, let us start with the most apt candidates for observation sentences. For any such typical candidate for observation sentence, it is very simple to conceive varied and prosaic situations where a verbal response to the sentence could not plausibly be directly and firmly correlated with the stimulations of the moment. Accordingly, these situations make all such sentences fail the proposed criteria. Here are some examples.

Take the most evident candidate, 'Red' or 'This is red'. Imagine a speaker, call her Lisa, in some dark place - say, a garage at night with no lights on - picking up an old plastic bottle with some liquid inside whose color is absolutely indiscernible in that light and uttering 'This is red' to her companion, call her Sam. Lisa has just found what she was looking for, the remainder of some red paint that she remembered she had saved there. The sentence fails to meet the criteria, the stimulation that prompted the speaker's utterance this time is not likely to prompt an equivalent response on a different occasion, nor is it likely to prompt an equivalent response from any other speaker witnessing the episode but ignorant of the fact that some red paint had been saved there. We cannot claim a firm and direct correlation between stimulations and responses for this sentence. This is not an observation sentence after all. There is more to the semantic value of this sentence than eliciting assent and dissent under certain stimulations. There is more to being competent in its use than being "conditioned simply [or outright,] to assert or assent to the sentence under some distinctive stimulation…" (Quine, 1993, 108).

In the opposite direction, we can just as easily imagine a situation where a typical presentation of red would fail to prompt assent to "Red". Imagine, for instance, that in a moment prior to their search in the garage, Lisa and Sam were looking for the red paint in Lisa's atelier. Sam, emerging from a messy closet holding a can of red paint - that is, a can with a commercial label featuring a red background sampling the color of the paint inside and the indication "Red Medium", and bearing some dried marks of spilled red paint - shows it to

Lisa saying “There you go. Red”. To which Lisa replies “No. That’s not it. That’s an old can I use to store varnish”. We see here that a typical presentation or stimulation is not sufficient. The first story had shown it to be also unnecessary. Patterns of stimulation seem never to be decisive.

Note that the point here is whether Quine manages to explain the use of certain sentences or expressions, not whether he manages to explain some instances of such uses. Quine’s general account of language depends on an adequate stimulation-response explanation of observation sentences, not of *occasional uses* of observation sentences. The notion in question is observation sentence, not *observational use of a sentence* or *observation utterance*. I should stress, in particular, that all the criteria put forward deal with and refer to sentences, not utterances. We do not even know what an observation utterance would be. In particular, Quine does not give us a behavioral criterion for their identification and there is no straightforward way to adapt the ones he does give. The most we can make of it is through a folk-psychological, vague and dubious - even by folk-psychology standards - notion of some utterance being produced in direct response to the stimulations without the speaker giving it some thought, or worrying if it fits well with what she already knows.

Accordingly, to show the failure of the account we need only to show that there are uses of the sentence that cannot be explained via some direct and firm connection to the stimulation; we do not need to prove that no use or utterance is.

Yet the latter is also likely the case. The examples or illustrations aim to show the former, but they also suggest - especially the last one - that the latter is also true. The fact that the speaker is in a position to use her mind and theory in a conspicuous fashion to *oppose* the stimuli when the occasion calls for it makes it not also clear that she must also be employing her mind and theory, albeit more discreetly, in *allowing* the stimuli to go unchallenged in the more typical cases? I believe it does point in that direction. It suggests that even in the most direct perceptual reports, mind and theory (or previous information) are always involved, more flagrantly when speakers - so to speak - *go against or without* the stimulus, but also in simply permitting speakers to *follow* it in the typical or regular way. However, as just explained, this is an argument I need not pursue here, and I will not explore it any further. To prove the unsuccess of Quine’s account, I need only show that there are no observation sentences, not that there are no observation utterances - whatever those are, if anything - and certainly not that there are no utterances very strongly determined by the speaker’s present

stimulations.

Returning to the examples, for every candidate observation sentence, it is clear that one can easily make up similarly convincing stories about situations where the sentence is not used to express the correspondent typical perceptual judgment.

Beyond first line candidates like 'Red' - and there are not so many of these - we get to sentences that Quine himself acknowledges to be less than purely observational.

> If 'Red' is somewhat less susceptible than 'Rabbit' to the influences of intrusive information, there are other sentences that are vastly more so. An example is 'Bachelor'. An informant's assent to it is prompted genuinely enough by the sight of a face, yet it draws mainly on stored information and none on the prompting stimulation except as needed for recognizing the bachelor friend concerned. As one says in the uncritical jargon of meaning, the trouble with 'Bachelor' is that its meaning transcends the looks of the prompting faces and concerns matters that can be known only through other channels. 'Rabbit' is a little this way, as witness papier-mâché counterfeits; 'Bachelor' much more so. (Quine, 1960, 37)

This seems like a first step in the direction of recognizing the problem I am trying to press, but a very shy one. Counterfeit rabbits should not be our biggest concern. Someone points to the oven and says 'That's a rabbit'; a child points to a mess of lines in his drawing and explains 'That's a rabbit'; John is telling Pete about a curious character from *Alice in Wonderland* describing it as a *cat that is* constantly *late,* and Pete naturally complains 'That's a rabbit'. Unlike with counterfeit rabbits, with these last cases there is not even any typical perception of rabbits prompting the verbal responses. Also, the question here is not one of vagueness around the edges of the set of prompting stimulations, nor are the situations presented extreme skeptical scenarios that one could dismiss as implausible and hardly to be taken into account when characterizing the uses of the sentence.

Other interesting examples of observation sentences proposed by Quine are mentioned in passages such as this:

> A chemist learns about compounds of copper in the course of his reading and experiments, and a physician learns about the facial symptoms of an overactive thyroid; in due course the one comes to recognize the presence of copper by a glance at the solution, and the other to recognize hyperthyroidism by a glance at the patient. The sentence 'There was copper in it' has become an observation sentence for the one, and 'He's a hyperthyroid' for the other. (Quine, 1995a, 43-4)

These pose flagrant problems to Quine, and the question here is not the one Quine identifies in the next passage, and promptly responds to.

> A second complaint is that what count as observations for the specialist often do not count as such for the layman. An answer to this objection is that the notion of observation sentence is relative to a linguistic community. If a sentence would qualify as an observation sentence for the scientist and not for the layman, it is couched in a technical sublanguage in which the layman is not a fluent communicant. (Quine, 1974, 41)

It is not the narrowness of the relevant community that worries me here. The problem is that even for the specialist no such sentence would satisfy the conditions set by Quine. It is not enough for a sentence to count as observational that its truth or falsity is *sometimes* open to view; it has to be so *always* - at least every time the sentence is pertinent, no matter whether true or false.

The individual criterion determines that for every stimulation, if it prompted assent once, it will prompt assent every time. Now, even if a physician, through long training, happens to acquire the ability to correctly apply the sentence mentioned based exclusively on the information afforded by a quick glance at the patient, surely she does not need to - and will not, in all likelihood - limit her uses of the sentence to such occasions. Sometimes she will be discussing a diagnosis with her colleagues, or she will need additional tests to decide about a less clear case, or she will be reading or writing about it. In such situations, the correspondent stimulation is not enough for prompting assent to the sentence, and they would fail to produce it again were they to be repeated.

The same happens with the social criterion. It requires that every time a member of the community assents or dissents to the sentence all others should agree with her. Let us assume that a very strong consensus prevails over such a diagnosis, but still the fact that only a fraction of such judgments is based on the moment's observation is sure to prevent the satisfaction of the social criterion. For the doctor who has been running all sorts of tests and studying the patient's medical history thoroughly, the stimulation prompting her final word, whatever that is, is very likely to produce a divergent response - most probably the simple withholding of opinion - on the part of another specialist.

6.1.2.5.1. COMPLEX OBSERVATIONALITY

Hylton, for one, defends that Quine can deal with some of these difficulties by making use of degrees of observationality (Hylton, 2007, 137). Allowing sentences to count as more or less observational seems like a promising way out of the above problems. All of the above sentences would count as more or less observational, and each violation of the criteria, instead of simply excluding them from a sharply delineated *observational class*, would simply slightly decrease its observationality.

Quine himself dwelled on this possibility but, in his last writings, he seems definitely to refuse it - with good reason, I think.

> In *Word and Object* (pp. 42-44) I consequently recognized degrees of observationality, and treated the intrusion of theory or collateral information as dilution of observationality. But in later writings, I held observationality as absolute, based on immediacy of assent, and then I accommodated the intrusion of theory by contrasting the holophrastic conditioning of the observation sentence to neural intake with the analytic relations of the component words to the rest of language. The sentence figures holophrastically both in the infant's first acquisition of it and in the scientist's immediate assent to it when testing a theory.
>
> My attention was turned anew to these matters by a recent letter from Lars Bergstrom adumbrating an empirical theory of truth. I ended up not adopting it, but it has prompted me to reflect further on the bipolarity of the holophrastic and analytic perspectives, as against the gradualism of observationality in *Word and Object*, or, as we might say, a gradualism of theoreticity.
>
> In conclusion, I retained the absolute notion of an observation sentence as simply an occasion sentence that commands the subject's immediate assent, however fallible and revisable. (Quine, 1996, 163)

I use only "he seems" because I find this passage - as well as others from the same period on the same topic - particularly equivocal. In the last paragraph Quine states his decision - to go with an "absolute notion of observation sentence" - but he still qualifies this decision showing that he is not willing to reject properties that do not combine so well with an absolute notion - *fallibility* and *revisability*.

In his last years, Quine returned with renewed interest to observation sentences and the problem of observation, where he finds a certain tension lurking. On the one hand, his theory and overall understanding of language and science require plain and simple observation sentences with their direct and firm correlation between verbal behavior and

stimulations. On the other hand, there is the pressure from the undeniable fact that all sentences are indeed structured, and that all assertions are contaminated by further theory, and do not respond exclusively to present stimulations.

Quine tried to accommodate both urgings by allowing observation sentences to have a double nature. Sometimes he talks about the *Janus-faced character* of observation sentences because of this double role, on the observational side, as mere responses to stimulations, and on the theoretical side, as structured and theoretically integrated assertions. I will not consider all the aspects of this ingenious position, but only a selected few that I will explore only enough to show that there really seems to be a problem with the lack of pure observational sentences.

Compare the last passage with this one from only three years earlier:

> We must recognize degrees of observationality. Assertion of the sentence or assent to it may be more or less delayed or hesitant. There may even be afterthoughts: 'Oh, it isn't a dog after all'. This sort of self-correction intrudes at the sophisticated stage where the child has come to appreciate component terms of observation sentences in their referential capacity. It is infection of observation by theory; the anti-epistemologists have a point here. But there are pure cases, and they prevail at the early stages of language acquisition. Observation sentences in this pristine purity are the child's port of entry to cognitive language, for it is just these which he can acquire without the aid of previously acquired language. (Quine, 1993, 108-9)

and with this one, from four years later:

> The speaker assents outright to "Rabbit" and retracts his assent later on discovering that what he had seen was a toy. In *Word and Object* (p. 44) I treated such intrusions of error as showing that the retracted sentence was not purely observational; theory had intruded. Theory had indeed, but I now see a better way of sorting matters out.
>
> Immediacy of assent is still my criterion of observationality, but emphatically and unequivocally so. The assent may be recanted, but its susceptibility to recantation becomes an independent dimension, *theoreticity,* for separate consideration.
>
> The range of neural intakes to which the speaker's assent to the sentence is keyed will of course be vague along the edges. The speaker may hesitate over "It's raining" in a fine mist, and over "That's a swan" in the startling presence of a black one. Vagueness of boundaries invests language at every turn, and I shall continue to take it in stride. But assent must be immediate when the stimulation is in the clean-cut range between vaguenesses. Immediate but fallible. (Quine, 2000b, 4)

It clearly results that Quine is struggling here for a way to conciliate both aspirations:

to do justice to the actual use of sentences by real speakers, and to preserve a solid link between language and theory and the world. If he were to let go of pure observation sentences, he would owe us an alternative account of that link. If not as *direct and firm* responses, in what alternative way, naturalistically describable, are verbal responses involving observation sentences responses to stimulation? I think it is because Quine is sensible to this point, and has no alternative answer, that he decides to preserve "an absolute notion of observation sentence" - he tries to hold on to the mechanisms of direct and firm response.

In the previous three quotations, from the oldest to the most recent, an unmitigated notion of observation sentence widens its scope of application as Quine delves deeper into the possibility of seeing the theoretical features of such sentences as an independent issue. Independence between observationality and theoreticity would mean no conflict between dimensions. It is because Quine becomes convinced that a sentence can be both observational and theoretical that he becomes more liberal with observationality. In the earlier passage, there is still the suggestion that most (if not all) sentences - outside immature utterances of children - are somewhat less than fully observational. Three years later, Quine is reclaiming an absolute notion of observation sentence and there is no indication, quite the contrary, that only very few expressions could hold the title. We are still left to wonder how this is possible, since he does not spare us the realistic disclaimer that such sentences are nonetheless revisable. In the last passage, he articulates something close to a solution proposal. Quine tries to overcome the incompatibility between observationality and theoreticity. However, I simply cannot see how this can be done. How can Quine hope to make independent those two properties - being observational and being theoretical - so as to allow them to live together in the same sentences? He just proposes the separation; he does not explain how it could be accomplished, or what it would consist of.

The apparent conflict seems most vivid for Quine in what he calls "the recanting of observation sentences" - in the last text that is where he focuses his attention. He tries to pass the idea that there are two distinct moments, a first one where the mechanisms of firm and direct response to stimulation still operate, and a subsequent second moment of theoretical critique – "It's raining", affirmed in full view of a rain-drenched window, is recanted on spotting the garden hose, and "Rabbit" on spotting the toy. (Quine, 2000b, 4) Perhaps there could be some consolation for the effective loss of observationality being pictured here in the

idea that, however theory manages to change the aspect of things, there is always an observational origin to responses involving observational sentences. But even that is hopeless. Theory affects the use of sentences in other ways besides commanding change of already produced verdicts and, as in the illustrations I presented above, responses to candidate observation sentences often do not have any interesting relation at all to occurring stimulations. Nothing resembling a first observational moment takes place in such instances. What is it for a sentence to be absolutely observational if that is compatible with its having uses not interestingly or decisively dependent on present stimulations? I could not find a good answer to this question in Quine and I do not expect one to be possible.

I believe we have no option but to conclude that there are no observation sentences, no sentences assent and dissent to those which are exclusively dependent on occurring patterns of stimulation. This should not really come as surprise. In Chapter 4, we noted the insufficiency of associatist accounts of word learning. Moreover, the idea of such automatic linguistic responses simply does not fit our natural experience of language and linguistic practices as we know that this is not how people use language. I think we must conclude that the irreducible theoreticity and intentionality of all sentences robs Quine of the observation sentences that he needs for his naturalistic account of language to work.

A minimally realistic appreciation of linguistic practices reveals that the same stimulations can prompt opposite responses to our best candidates for observation sentences, and that sometimes the responses have really nothing to do with the superficial features of the presentation. Whether or not there was ever a moment - phylogenetically or ontogenetically speaking - where individuals' responses to sentences were directly and firmly dependent on occurrent patterns of stimulation[123], that is not what normal language and language use is like, and at no moment do speakers return to these primitive transactions.

This negative conclusion only strengthens the natural, folk-psychological view that the crucial dependencies are not between linguistic responses and sensory stimulations, but between linguistic responses and speakers' mental states. When speakers believe they are being presented with a rabbit, or something red, they will tend to assent to "Rabbit" or "Red", provided they also bear the adequate conative states.

[123] Once again let me remind you that in Chapter 4 we saw that, from the early beginning, associationism and the idea of simple responses to stimuli - be they distally or proximally understood - do not explain much of what is going on in language acquisition.

6.1.3. EXPLAINING PERCEPTUAL SIMILARITY

In the previous sections, I looked for Quine's behavioristic criterion for the identification of observation sentences and perceptual similarity among stimulatory episodes prompting similar verbal behavior. I found criteria that I believe can be adopted for the identification of sentences' verbal responses which are directly and firmly keyed to occurrent stimulations, but defended that they are satisfied by no sentence. Let us ignore, for now, this crucial unavailability[124], and move on to consider for a moment the next step in line, as anticipated above, in a naturalistic account of observation sentences and perceptual similarity - the *actual explanation* part.

We must emphasize that a behavioristic criterion cannot be confused with a real explanation. If what we want is a naturalistic explanation of linguistic behavior based on stimulation properties, the signaling of the properties in question by reference to the same behavior it was supposed to explain cannot amount to any kind of progress in the understanding of that behavior. As an explanation it would be just as circular as it can get: *this is the type of behavior one will verify each time the sensory stimulations are such as to bring about this kind of behavior.*[125] Even if one were to obtain a clear behavioral test for the precise determination of the cases where perceptual similarity has been responsible for the production of such behavior, one would not have advanced a step towards an explanation of such behavior. This is a fundamental point that must not be obscured by all the difficulties surrounding the previous questions.

Up to now, all that we saw Quine doing was to posit a naturalistic property, perceptual similarity – a physical analogue for a mentalistic property about which we seem to possess some more or less clearly defined intuitions – and to strive for a behavioral test capable of identifying its presence. Quine, at times, is perfectly clear in this respect,

> Mental states do not reduce to behavior, nor are they explained by behavior. They are explained by neurology, when they are explained. But their behavioral adjuncts serve to specify them

[124] There are still similarities between perceptual circumstances, and recognition of such similarities, even if no sentence is such that its use is solely determined on the basis of such recognition. Thus, Quine's ideas on perceptual similarity can be assessed independently of his ideas about the role of this relation in the determination of linguistic behavior.

[125] Hylton, 2007, 119, makes explicit the same conclusion about the explanatory nullity of finding a behavioristic criterion.

> objectively. When we talk of mental states or events subject to behavioral criteria, we can rest assured that we are not just bandying words; there is a physical fact of the matter, a fact ultimately of elementary physical states. (Quine, 1978, 167)

A behavioral test is not fit as an explanation; it is only something provisional, a warrant that *there is something there* worth exploring.

Quine has no neurological or physical theory to offer, but still one might be tempted to read passages such as the following as delivering some sort of positive proposal concerning the second part of the exploration.

> … most of the [sensory] receptors triggered on any occasion are perceptually ineffective. What matter are the salient ones. We can now proceed to narrow our sights to these, for salience is definable in terms of receptual and perceptual similarity. The receptors that make for salience within a global stimulus are the ones that the stimulus shares with other global stimuli to which it is perceptually similar but receptually dissimilar**.** The salient part is the part by virtue of which the global stimulus is perceptually similar to others despite divergence of other parts. (Quine, 1995a, 17-18)

This passage could be read as suggesting something in the way of an explanation for perceptual similarity, the very most simplistic one we could imagine: a stimulation is perceptually similar to other stimulations because there are some "receptors … that the stimulus shares with [those] other global stimuli" - the sharing of receptors could likely be understood as some nervous terminals being activated in the same way in each episode. In other words, this would amount to claiming that there is perceptual similarity because there is partial receptual identity.

Such a proposal would be so naively and flagrantly wrong that, despite the arguably supportive passage, it seems unacceptably uncharitable to attribute it to Quine. Take the short example already mentioned. Compare a cat seen at night in an alley and a cat lying in the grass at noon. Should we expect some of the observer's photoreceptors to be in the same state on each occasion and for that to be the determinant factor causing her to perceive them as perceptually similar? Blatantly, that cannot be the case. Photoreceptors, of course, respond to light. All sorts of luminously different occasions prompt acceptance of "There goes a cat". Each time, the light striking the speaker's retina will excite the sensory receptors there in completely different ways. Quine cannot fail to be aware of this.

Despite the occasional concrete tone of his talk about sharing receptors and the like in passages such as the previous one, the fairest conclusion to take, I am convinced, is that

Quine has no positive contribution at all to offer for an explanation of perceptual similarity. I think we should read him as merely persisting in the voicing of his conviction that there must be something in common between the relevant stimulation patterns - whatever that may be, and however that can be accounted for in naturalistic terms. Assuming that that is the case, my last comments will be equally brief and superficial.

On the one hand, I think we should recognize that *a partial* confirmation of Quine's expectations is indeed almost at hand. In the last few decades, there has been prodigious progress made in the sciences of perception and with the science and technology of automatic image interpretation. Such progress has made it abundantly patent that the possibility of two stimuli having something in common does not need to be realized in so crude and dull a fashion as considered above. Two images - let us think of them as two compositions of juxtaposed colored points - completely different on superficial inspection, incoincident point by point in whatever portion one chooses to compare, can still be shown to be identical under some particular aspect, and have that identity objectively confirmed through mathematical processes of interpretation and analysis. This is what happens when two very different pictures are classified under the same category - for instance, as a human face - by some image recognition software. The equivalent, to some degree, should also be true of stimulations of retinal receptors, that is, they must bear analog similarities *beneath their surface.*

On the other hand, such advances still leave us very far from a realistic understanding of the processes taking place in the kind of situations Quine has in mind and of his notion of *perceptual similarity*. I suspect the notion to be too heterogeneous and loose, bearing the marks of a high-level notion, fit for general and synoptic explanations, human understanding and mundane navigation, but not for the kind of rigorous and detailed science that Quine favors. Notice, for instance, that cases such as those mentioned by Quine - the perceptual recognition of red, of rabbits, of rain - are strongly determined by rich and multifarious factors such as *attention*[126] with subsidiary issues including context, interests, background, and so on. If we remove these factors from the picture, it is very doubtful that we end up with a notion of perceptual similarity that suits Quine's purpose of accounting for the classifying together, by an agent, of perceptual episodes. However, the idea of understanding and

[126] For a brief development of what I have in mind and further references, see Braisby & Gellatly, 2012, chapter 2, and Foley & Matlin, 2009, pp. 133-137.

integrating the workings of such factors as attention so as to provide a complete enough neural level account of perceptual similarity seems so remote and vague as to be uncertain. Despite Quine's efforts, *perceptual similarity* still looks like a notion simply borrowed from folk psychology. We grasp this notion quite easily, in commonsensical and intuitive terms. It plays a role in a certain type of account, and is fit to respond to particular types of concerns but not others. Orthodoxy in philosophy of science[127] tells us not to expect a direct and clean reduction of such notions into any lower level science or description.

6.1.4. THE FAILURE OF QUINE'S NATURALISTIC ACCOUNT

At this point I think it is fair to say that the basis and core of Quine's proposal for a naturalistic account of language is seriously undermined by the problems presented and developed in this chapter. First - and that is my main argument here - there are no observation sentences, that is, no sentences whose use is explainable as firm and direct responses to stimulations, nor to perceptual situations, speaking more generally and neutrally. Second, perceptual similarity is not explained and looks just as irreducible to (Quinean-) naturalistic terms as before the little Quine had to say on the subject.

Hylton, at the end of his long book, synthesizes Quine's project and goals and articulates a corresponding success test.

> ... the criteria for the success of Quine's genetic project are clear-cut. At least as Quine presents that project, it is a more or less straightforward scientific issue: can we give a purely naturalistic account of a child's acquisition of language and knowledge? (Of particular importance here is the fact that a purely naturalistic account would not presuppose terms or principles which Quine would dismiss as mentalistic.) Or better, perhaps: do we have a reason to think that such an account is in principle available, even if we are not, and may never be, in a position to give it in detail? The project would, presumably, succeed if we had a clear affirmative answer to this question. Quine claims to have given a sketch of an account of the required kind, a sketch that makes it plausible that a full account is in principle available. … So Quine must be understood as aiming to make it plausible that it is, in principle, possible to give a purely naturalistic account of the phenomena of the acquisition of knowledge and of cognitive language, when those phenomena are described in what he would take to be a naturalistic fashion. (Hylton, 2007, 363-4)

[127] See Hardcastle, 2009, Thagard, 2013, and McCauley, 2007.

Hylton himself offers no verdict, he observes that "[o]pinions may differ as to whether [Quine] has in fact succeeded in this task" (ibid.) and deliberately abstains from producing his.

There are more openly favorable pronouncements. Gary Kemp, for one, has very recently professed his persuasion that Quine's proposals constitute indeed a positive contribution to the naturalistic understanding of language, meaning and communication.

> the vast majority of his published output, beginning with *Word and Object* in 1960, has been devoted to an ambitious positive task, what Peter Hylton (2007) calls the 'genetic project': the project is to show that it is at least in principle possible to give an account of how knowledge could be acquired, staying within the confines of naturalism. And central to such a project will be to account for the possession of cognitive language, the medium of theories. Moreover, assurance that a successful outcome to the genetic project is possible is vital to maintaining naturalism, or at least it is vital to maintaining it in a healthy state: if it's not successful—that is, if it proves inevitable that we should have to use such terms as 'understanding', 'grasping a proposition', 'meaning', 'belief ', and so on, in un-explicated ways, ways that go beyond the strict discipline enforced by naturalism— then there would be an apparent counterexample to naturalism's signature claim: human language. …
>
> … I will put the case that Quine's naturalized account of language can do the trick without any use of semantical concepts, as traditionally conceived, at all. (Kemp, 2012, 11-2)

This is stated in the introduction. Afterwards, throughout the book, he attempts what he promises, i.e. to "put the case that Quine's naturalized account can do the trick". He sympathetically and mostly uncritically recapitulates Quine's proposals, without seeming to notice any serious problem or insufficiency - in particular without seeming to notice the difficulties I pointed to in Quine's account of observation sentences and perceptual similarity.

I, of course, disagree. Has Quine shown how to account for language, meaning and communication in (Quinean-)naturalistic terms? Has he, in some way, rendered its possibility more likely? Has he, somehow, contributed to strengthen the belief that it can be done? Again, based on the preceding arguments, I believe the answer to all these questions must be negative.

I am not disputing the obvious claims involved, such as that the stimulations of our sensory terminals exhaust the information input into the organism, or that all learning, of language or world, must start with what is open to view. Neither do I reject the virtual consensus over the belief that our linguistic competence as well as all our mental states must have a biological, neurological and physical realization. My point is that beyond an appeal to

these generic and almost universal convictions, Quine produces little to nothing more in favor of the possibility of a (Quinean-)naturalistic account. As I see it, and have argued for, Quine adds nothing significant to the realization or plausibility of such a project

6.2. TRANSLATION NATURALIZED

Without prejudice to my negative conclusions on the adequacy of Quine's proposals for a naturalistic account of language use and acquisition, I still want to explore a few more aspects of his take on the topic, in particular, his proposals for naturalistic methods of translation or interpretation. These naturalistic methods could constitute themselves as alternatives to interpretation via rationalization and, thus, as a challenge to the thesis I am defending here that interpretation requires rationalization.

Let us see, then, what Quine offers or suggests. According to Quine what is one to do in order to translate, or, more generally, in order to acquire the ability to understand the linguistic utterances of another speaker? I think we can find or support several answers to this question in Quine's work.

6.2.1. JUST BE NATURAL

The first plan - we must not omit it, being arguably the most congenial to a (Quinean-)naturalistic conception - is just to be a natural and physical object in a natural and physical world and to react in accordance to the fundamental laws of nature to the impacts of various forces and molecules. In a sense, the recipe is probably correct; this is a naturalistic form of translation, even if not really a method or plan of action, and we can call it *the naturalistic course* of interpretation or language acquisition. Does it refute my thesis, that interpretation requires rationalization?

For now, there is certainly nothing close to a complete (Quinean-)naturalistic understanding of these processes. Quine himself, I defended, fails to contribute to any progress on this front. But I am definitely not rejecting the possibility of future developments in science that could bring us to that point. Perhaps someday we will able to describe in (Quinean-)naturalistic ways that completely dispense with any appeal to rationalization the processes that bring about an interpreter's ability to linguistically understand some speaker. I

prefer to risk no stance on the likelihood of such a possibility.

Even if that were to arrive, it would not fatally imply any dispensation of rationalization. We are dealing here with different levels of description, and with different perspectives of the same events. Even if interpretation someday came to be adequately understood at a neural level, and from a third-person perspective, this would not, in any obvious way, liberate interpreters from the need to engage in their old-fashioned investigations and methods. A rationalizing inquiry could still remain the indispensable psychological correlate of those neural events.

On the other hand, it could also happen that such hypothetical breakthroughs came actually to inform and reform interpreters' conduct in such a way as to really refute the thesis here defended. That is, it is certainly a possibility that scientific progress could, additionally, somehow come to displace ordinary interpreters out of their folk ways and into improved interpretation practices of some sort. But this is an even farther prospect than that of achieving an external, third-person, naturalistic understanding, and it poses, at present, no substantial threat.

6.2.2. THE HYBRID METHOD

There is a second line of approach, one which seems to be located more or less halfway between a purely (Quinean-)naturalistic method such as the one just described and a fully commonsensical one, involving rationality, rationalization and all the proscribed psychological notions. Quine presents this hybrid method in the second chapter of *Word and Object*, where it is to be employed in the conceptual experience of *Radical Interpretation.* That is where Davidson, Quine's student, picks his inspiration for *Radical Interpretation.* In its goal and general lines, the task attributed to the radical translator is identical to that attributed to the radical interpreter. Both must develop the capacity to understand the words of another person - be it via a translation manual or via a meaning-theory - starting from an initial position of total ignorance of her language. The difference, naturally, is with the methods to be employed by each of them, as well as the philosophical background supporting those methodological options.

It must be stressed that Quine's interest is not with translation itself, nor with a realistic design for an actual translation method. He uses the experience and the naturalized

conception of translation to support his rejection of the traditional view of meaning and understanding, and, in particular, his indeterminacy claims. He is trying to show that there are no good naturalistic grounds for the employment of our traditional notions, that we cannot naturalistically individuate entities such as meanings and propositions in any way resembling what we would need to ground and confirm the natural and pre-theoretical talk about meanings, sameness of meaning, communication as meaning exchanges, and so on. Nonetheless, it is still a hypothetical alternative to our natural, essentially mentalistic method of going about learning other people's words that Quine coins in the experience - one that could, if proven viable, challenge the indispensability of rationalization and psychology for interpretation.

Quine's simple account of linguistic competence and behavior involving observation sentences suggests an equally simple translation method for those sentences[128]. The radical translator must "correlate the native's observation sentences with his own by stimulus meaning" (Quine, 1960, 42). He would only need to look for pairs of observation sentences - one from the known, *target language*, the other from the unknown, *object language* - each associated with the same, or sufficiently similar, stimulus meanings – "the imagined equating of 'Gavagai' and 'Rabbit' can now be stated thus: they have the same stimulus meaning." (ibid., 29) If the stimulus meanings match, the sentences translate each other, or at least they bear a relation to each other that is the closest to translation that can be made naturalistic sense of. This was Quine's first and long-espoused view on the topic, prior to his change of spirit, already explained above, concerning the idea of shared stimulations.

Notice that this method appears to dispense with any involvement of mindreading and rationalization. The other speaker is to be seen as a black box devoid of mentality, simply reacting in regular neurological, biological and physical ways to its environment.

6.2.2.1. NO ROOM FOR THE PRINCIPLE OF CHARITY IN THE TRANSLATION OF OBSERVATION SENTENCES

Often, Quine's and Davidson's takes on interpretation are made clearer when contrasted with each other. A particularly interesting divergence has to do with Quine's

[128] Once again, I will focus only on these sentences as the translation of the rest of the language would be dependent upon theirs.

conception of the important difference between types of sentences - observation sentences versus the rest - a difference that is reflected in his recognition of two distinct stages in the translation enterprise. What is at stake here is the epistemic status and quality of the translation hypothesis in each of those different moments.

> Sentences translatable outright, translatable by independent evidence of stimulatory occasions, are sparse and must woefully under-determine the analytical hypotheses on which the translation of all further sentences depends. (Quine, 1960, 65)

The passage contrasts two different epistemic moments. First there is an initial moment of translations solidly grounded on empirical evidence, concerning only observation sentences. Each particular translation proposal about a certain native sentence will constitute "a genuine hypothesis from sample observations, though possibly wrong. 'Gavagai' and 'There's a rabbit' have stimulus meanings for the two speakers, and these are roughly the same or significantly different, whether we guess right or not." (ibid., 66). There is no need here for rationalization, mindreading, and the principle of charity. The principle of charity was first introduced by Quine into the question of a radical investigation of an unknown language. Davidson borrows it from Quine, but Quine uses it, parsimoniously, only in later stages of the process.

Things are different for the non-observation sentences, let us call them *theoretical sentences*. Their translation requires what Quine calls *analytical hypotheses*. Having settled the translation of observation sentences and of the truth functions - together with a couple of other facts about *stimulus synonymy*[129] that we do not have to worry about here - the radical translator will then need to employ some creative work. We can say - very crudely but fitting our purpose - that the translator must now design a way to segment the translated sentences, and corresponding translations, into pieces out of which it should be possible to reconstruct not only the sentences already translated but also all other sentences, and their corresponding translations. These are *the analytical hypotheses*. They do not have empirical tests of their own, they extend "beyond the zone where independent evidence for translation is possible" (ibid., 65). The only empirical constraint determining their invention - a quite loose one, according to Quine - is that they must warrant the same translations for occasion sentences and logical connectives that were already established on previous, more solid grounds.

[129] See Quine, 1960, 62.

This marked difference between translation stages, with its affirmation of a first moment of secure translation based on the purely objective and measurable features of the situations, appealing to correlations between stimulations and responses that are supposed to be identifiable on rigorous and unequivocal terms without the need to employ the filter of any mind, has no replica in the Davidsonian setting of Radical Interpretation. As Davidson explains in the following passage, this is what determines the need for a much stronger appeal to the principle of charity.

> I would extend the principle of charity to favor interpretations that as far as possible preserve truth: I think it makes for mutual understanding, and hence for better interpretation, to interpret what the speaker accepts as true as true when we can. In this matter, I have less choice than Quine, because I do not see how to draw the line between observation sentences and theoretical sentences at the start. There are several reasons for this, but the one most relevant to the present topic is that this distinction is ultimately based on an epistemological consideration of a sort I have renounced: observation sentences are directly based on something like sensation—patterns of sensory stimulation—and this is an idea I have been urging leads to skepticism. Without the direct tie to sensation or stimulation, the distinction between observation sentences and others can't be drawn on epistemologically significant grounds. (Davidson, 1983, 149)

Or, much more synthetically stated in the preface to *Inquiries into Truth and Interpretation,*

> Quine applies the principle primarily to the interpretation of the logical constants.
>
> Because I find I cannot use Quine's notion of stimulus meaning as a basis for interpreting certain sentences, I apply the Principle of Charity across the board. (Davidson, 1984d, xvii)

It is by now perfectly clear why we do not find in Davidson a clearly defined category of sentences, analogous to Quine's observation sentences, with which to start Radical Interpretation. One can, as I have done in Part I, point to a set of sentences - I called them *perception sentences* - whose content is, in some sense, more easily *readable* from the occasions on which they are held true - or, in Davidson's terms, "whose causes of assent come and go with observable circumstances and those a speaker clings to through change remains" (Davidson, 1983, 149). However, unlike what must be the case with Quine's observation sentences, the Davidsonian set of perception sentences is essentially vague and fluid.

In Davidson, no very significant or sharp difference is established between types of

sentences, their respective interpretation methods, or between the nature or epistemic quality of the interpretative hypothesis. All we have in Davidson is a smooth gradation, eventually shifting from context to context - the same sentence that prompts a perceptual response on a certain occasion can prompt a non-perceptual one in a different moment. Moreover, what determines if a certain sentence counts as perceptual or not, for a certain speaker, on a certain occasion, is not - in contrast with Quine's story - some objective and external property testable previously to, or, at least, independently of, interpretation itself. The contrary is the case. Determination comes from interpretation; it already requires interpretation. The perceptual character of a certain sentence is constituted only by reference to the speaker's established system of meanings and beliefs. Only when all that is in place can one tell what is and what is not *open to view*.

6.2.2.2. EXPLANATORY LACUNAS DO NOT DEFEAT THE HYBRID METHOD

But Davidson is not the point here, Quine is, and his potential naturalistic method of translation through matching of stimulus meanings - the hybrid method. What I want to consider now is whether or not, and to what degree, the failure of Quine's naturalistic account of language defended earlier in this chapter compromises the adequacy of this method of translation. In anticipation let me declare that I do not think the detected failure of Quine's explanatory project is enough to jeopardize the hybrid method of translation. Let us see why.

Quine's explanatory project, his foundational semantics, is entirely dependent on the possibility of establishing firm and direct correlations between stimulation patterns and speakers' linguistic responses. The stimulations would simultaneously cause and explain speakers' responses. Quine's strategy is to use this sort of direct correlation to explain linguistic behavior involving the most basic class of sentences, observation sentences, and to invoke inferential relations between those sentences and all others to explain the remaining uses.

As explained above, the project fails first of all because there are no observation sentences. Any explanation based exclusively on the idea of *response to stimulations* will be incomplete because no sentence is such that its uses are directly and firmly determined by present stimulations. Mind and theory always intrude; we never get direct responses. Take a case where we can plausibly conceive that we are close to achieving that direct link - the use

of "Red". Imagine that we were indeed to discover a certain regularity φ among most of the stimulations preceding speakers' assent to "Red". We can say, approximating Quine's usage, that stimulations exhibiting φ are perceptually similar among themselves. Imagine now that φ is found in 90% of all stimulations preceding assents to "Red", but is not there in the remaining 10%. Let us establish that there is no mistake here, that in those 10% of cases the assent was correct and genuine but simply had really nothing directly to do with the speakers' perceptual states - remember the story, presented above, about finding the can of paint at night in the garage. Now, even in the event of arriving at the discovery of such a frequent regularity as φ, the fact of these inevitable exceptions would betray the incompleteness of the type of *stimulation-response* explanation employed by Quine. The unaccounted for 10% imply that the hypothetically detected regularity does not exhaust all that is relevant in the prompting of responses to the observation sentence "Red". In other words, there is more to that sentence's semantic value than being keyed to certain stimulatory situations, and there is more to being competent in its use than being conditioned to assent and dissent to the sentence in certain types of stimulatory situations.

However, the failure on the *explanatory side* does not imply a failure on the *translation/interpretation side*. On the translation side the incompleteness might not be so damaging - a merely frequent correlation could be enough for the purpose at hand. Even without the grasp of all the relevant factors, it could plausibly happen that the hypothetical detection of a merely frequent regularity is sufficient to establish the correspondence aimed at between expressions of object and target language. It could even work for yet less frequent correlations.

Here is another primitive illustration to render more clearly what I have in mind. Imagine that a certain stimulatory regularity β is found preceding assent of around 25% of speakers of Jungle[130] to "Gavagai". Imagine further that β is also found preceding English speakers' assent to "Rabbit" in a close enough percentage of times. Imagine, finally, that no other English sentence presents a credible challenge to "Rabbit" as a correspondent in English to "Gavagai" - imagine, that is, that for no other English expression is it the case that regularity β is found preceding, let us say, *5% or more* of speakers' assents to it. In such a case, and under certain assumptions, it would not be unreasonable to use such discoveries to

[130] "Jungle" is sometimes offered as the name of the imagined remote language.

support that “Rabbit” translates “Gavagai”.

The lesson here is that even if one does not grasp enough of what is going on to constitute an alternative naturalistic account of speakers’ use of certain sentences, one might still have enough to engage in Quine’s hybrid method of translation. Despite the failure of the explanatory project, these ideas could still be defended as containing at least the germ of a naturalistic method of translation.

In response to this view, in the next section I will defend that this hybrid strategy, even if it were to work, is still built over a groundwork of mindreading and rationalization. The (Quinean-)naturalistic step, the matching of stimulus meanings, can only take place after a number of non-naturalistic pre-arrangements.

6.2.2.3. THE UNDERLYING RATIONALIZATION

I distinguish three ways, or moments, through which rationalization sneaks into the translation. First, note that the detection of some hypothetical affinity between sentences of different languages - that is, the discovery of some hypothetical common frequent correlation between a certain stimulatory regularity and speakers’ behavioral responses to each type of sound or inscription - by itself amounts to no translation. Only against the backdrop of a certain conception of language and of linguistic practices can the noted affinities begin to reveal anything. The translator must realize or assume that the foreign speaker is using her sentence just like the translator uses, or might use, his, and only from this can he conclude that the sentences must be approximate translations of each other. However, this assumption or realization is loaded with mindreading and rationalization: it involves seeing the speaker as a participant in the general business of linguistic interchange and communication, not to mention thinking and action, and requires being able to interpret enough to detect which moves of hers count as linguistic responses to the sentences in question. All this is at least implicit in Quine’s depiction of the experience. But there is more beyond these general assumptions and the ability to recognize behavior as uninterpreted linguistic responses.

Second, we must stress that Quine’s experiment makes a crucial, if only tacit, appeal to a yet more specific understanding of the context speaker and translator are involved in, an understanding that must be shared by both, and without which it would become very implausible that the hybrid method could be used with any success at all. What I mean is that

it must be common knowledge between speaker and translator that the context they are engaged in is one of translation or interpretation from scratch. In Quine's description of the affair it is clear that translator and speaker must cooperate to some degree and that such requires that both realize that what is being endeavored is the transmission-learning of the speaker's language. Note, for a start, that the speaker must be aware of the speaker's initial ignorance of her language, she must realize that he cannot understand her words and adapt her linguistic behavior accordingly. But the translator himself is expecting that sort of collaboration and tries to make the most of it. Consider the role of the translator's inquiry as an example elucidative of the indispensability of this mutual understanding. In Quine story, the translator's questions have a fulcral role since he must test his first translation hypotheses as he keeps gathering them. But these are not regular questions, and for them to work they must be understood in their peculiarity more as questions concerning language than as genuine questions. No one utters "Rabbit?" as a literal question when it is more than obvious that the animal the utterer is pointing to is a rabbit. Unless the speaker realizes that the speaker is actually inquiring about the language, the translator will not be getting adequate answers of the type he needs to progress in his investigation. My point, of course, is that speaker and translator will have to involve rationalization to be able to reach this necessary common knowledge of their specific engagement.

Third, we can detect still more mindreading and rationalization involved in the affair when we attend to what is involved in the emerging and testing of specific hypotheses of translation. The translator does not start from a neutral examination of the scene of assent or dissent to form his hypothesis; he starts with some unspecified but rich background knowledge, a set of assumptions. He knows what is plausible or likely and what is not, and that is what guides the inquiry. Even when he flirts with *radical questioning* - rabbits or temporal segments of rabbits? - common sense is still the flagrant route which he has to force himself away from, and from which he cannot completely break free. Common sense is the only source of plausible progress but also the fuel for these somewhat fabricated and fruitless excursus. Afterwards, in putting these hypotheses to the test, the translator will make use of diverse means to increase clarity and focus. He will point to things, mimic them, choose clear targets, avoid noise and ambiguity, and so on. All these efforts - the anticipation of probable meanings and contents, the focus on the speaker's perspective, on their experience of the scene and on some privileged stimuli, and his efforts to detect and guide the speaker's attention - constitute new forms of mindreading and rationalization.

We find in Quine what can be seen as the rudiments for a naturalization of interpretation. What I have done in this section was to highlight the non-(Quinean-)naturalistic nature of certain essential steps involved in the alternative method. I can accept that there is a grain of plausibility to the story of the radical translator as it is told, most especially, in chapter two of *Word and Object*. However, that is due, according to my diagnosis, to an intense but implicit involvement of mindreading and rationalization in the affair. Take away those elements and the experiment, along with the viability of a method of translation based on the matching of stimulus-meanings, plunges into the deepest unlikelihood.

It is not certain how strong a case in favor of the possibility of a naturalistic method of interpretation Quine thinks the conjectured viability of this hybrid method would make, but I believe that the explorations in this section show that it is not a very strong one. I conclude that no serious threat to my claim comes from here.

6.2.3. TRANSLATION THROUGH EMPATHY

This third method, translation through empathy, is offered by Quine as a new and very significant variation on the initial method of the radical translator. Quine's later dissatisfaction with some of the ideas he had employed in his *Word and Object* methodology dictated these changes.

As already introduced above[131], by the time of *The Pursuit of Truth,* Quine is dropping an important part of the notion of *stimulus meaning* and revising its role in translation. He rejects his initial conception of sameness or closeness of stimulus meaning on account of stimulations being a private and idiosyncratic affair.

> … an event of stimulation (...) is the activation of some subset of the subject's sensory receptors. Since the linguist and his informant share no receptors, how can they be said to share a stimulation? We might say rather that they undergo similar stimulation, but that would assume still an approximate homology of nerve endings from one individual to another. Surely such anatomical minutiae ought not to matter here. (...)
>
> The view that I've come to, regarding intersubjective likeness of stimulation, is rather that we can simply do without it. (Quine, 1990, 41-2)

[131] See 6.1.2.2. Social Criterion B.

Quine lets go of any idea of intersubjective identity or likeness of stimulations. He withdraws to a less demanding notion of *harmony* (Quine, 1995a, 20-1) in the way different people react to similar circumstances. This weaker notion seems enough for the sort of agreement needed to account for communication, language learning and translation. The "coordination of behavior across the tribe" (Quine, 1995a, 20) requires no sharing of *the* contents of *the beetle box*[132] nor of its physical correlates. He explains this harmony as the joint product of natural selection and common culture.

With this, Quine can no longer explain interlanguage synonymy in terms of identity or closeness of stimulus meanings as he did in *Word and Object*. Translation can no longer be a matter of finding a sentence in the linguist's repertoire, assent and dissent to which is prompted by the same stimulations that would prompt the native's assent and dissent to the sentence being translated.

To circumvent this difficulty, Quine gave prominence to the notion of empathy and, with it, blurred any distinction between the resulting method and a fully intentional one, involving rationalization and all the proscribed psychological notions.

From here on, he ceased resisting this kind of psychological account, and ran ever closer to an exclusively commonsensical view of the language learning process.

> Empathy dominates the learning of language, both by child and by field linguist. In the child's case it is the parent's empathy. The parent assesses the appropriateness of the child's observation sentence by noting the child's orientation and how the scene would look from there. In the field linguist's case it is empathy on his own part when he makes his first conjecture about 'Gavagai' on the strength of the native's utterance and orientation, and again when he queries 'Gavagai' for the native's assent in a promising subsequent situation. (Quine, 1990, 42)[133]

He now talks freely of the translator's dependence "early and late on psychological conjectures as to what the native is likely to believe" and declares that "[p]ractical psychology is what sustains our radical translator all along the way, and the method of his psychology is empathy: he imagines himself in the native's situation as best he can" (ibid., 46). All hints of an alternative, naturalistically-based active approach to the understanding of someone else's words have vanished. In the end, what remains is but the ordinary method of

[132] Allusion to Wittgenstein's *Philosophical Investigations*, §293.
[133] See also Quine, 1995a, 80 and 89.

translation or interpretation, in whose description and understanding one does not advance far beyond what the folk can make of it. We seem to have available but one type of account of what takes place in understanding, and mentalistic notions and common sense psychology occur there essentially. Quine seems no longer reticent about the fact that he has nothing better or different to offer.

Note, finally, that giving up on naturalized translation or interpretation is not the same as giving up on a naturalistic account of language. We must understand this to make sense of the coexistence in the same texts of psychological accounts of empathetic translation and talk about perceptual similarity between global stimuli explaining identical linguistic responses. Quine never abandons his conviction that there is a (Quinean-)naturalistic account of language to be sought and told; he just puts on hold the idea that one must be able to extract a method for naturalized translation or interpretation from such an account.

6.2.4. RATIONALIZATION PREVAILS

There is no substantial and thoroughly naturalistic alternative method of translation or interpretation to be found in, or inspired by, Quine's work. We have considered three different takes on the task.

I started by admitting the likelihood of there being alternative descriptions, produced at a neural, chemical or physical level, of the same events that constitute the common process of interpretation that linguistic agents go through. Then I argued that such descriptions - descriptions of the *naturalistic course* of interpretation - would fall short of refuting the claim that interpretation requires rationalization. They would depict the rationalizing processes from a different angle but they would not support the idea that there can be interpretation without rationalization.

Then I observed that the matching of stimulus-meanings promised a very simple - in theory, not in practice - alternative to a rationalizing method. I revealed several ways in which rationalization is needed to set the stage for any operation of the sort. The hypothetical viability of such a *hybrid method*, I concluded, poses no threat to my general thesis.

Lastly, I reported Quine's full resignation to a commonsensical and rationalizing methodology in interpretation. I stressed that a concession on the *interpretation methods* side

does not need to be accompanied - and indeed it is not - by any equivalent surrender on the *explanation* side.

Although I focused exclusively on Quine's work and proposals, I believe that some of the reasonings and arguments employed here can also have application elsewhere. There should be other proposals for naturalistic methods of translation or interpretation on offer, presenting apparent challenges to the central claim of this dissertation, but that can be neutralized with like responses to those produced in dealing with the *naturalistic course* and with the *hybrid method*.

Part II.b

Rationalization and Convention

Introduction to Part II.b

In this Part II.b, I consider the role of conventions in linguistic communication, and the relation of these institutions to interpreters' rationalizing practices. Two central questions emerge here: "Are conventions sufficient for linguistic communication?" or, in other words, "Can conventions free interpreters from the need to rationalize their interlocutors' utterances, thus compromising the thesis defended in this dissertation?", and "Are conventions unnecessary for linguistic communication?" or, in other words, "Is it the case, on the other hand, that conventions are no more than a helpful convenience that could be dispensed with by relying exclusively on interpreters' ingenuity and rationalizing abilities?" To both questions I give a negative answer. Only the first one is decisive to my thesis. The inquiry into the second point is meant to allow us a more complete and fairer picture of the intricate relations between social institutions like conventions, whose function, at least in part, is to preserve the linguistic uses of the community, and rationalization, the interpreter's method for making his way into such established uses, but also for dealing with creation and change in language. It will result clearly, I expect, that far from being in tension, the two strategies or devices neatly complement each other.

In Chapter 7, I address a form of Radical Anti-Conventionalism - one arguably maintained by Davidson - that affirms both the general insufficiency and the general dispensability of conventions for linguistic communication. My main focus here are the arguments against necessity, and I conclude that they are ineffective in supporting a strong and general claim of the dispensability of conventions in linguistic communication. Speakers will often manage to communicate linguistically without sharing conventions determining the use of some of the expressions that they, nonetheless, competently exchange. What I cannot find is a persuasive argument supporting any form of extrapolation from this local level claim to a more general level one - that is, any form of extrapolation from claims concerning the interpretability, without conventions, of occasional expressions, to claims concerning longer stretches, or even the totality, of discourse.

In Chapter 8, I address a form of Radical Conventionalism advocated, very recently,

by Ernest Lepore and Matthew Stone (L&S) in their 2015 book, *Imagination and Convention.* L&S affirm that all linguistic meaning belongs on the conventional side of the divide that they propose as the fundamental one with which to address uses of language. Outside convention there are only the imaginative forms of engagement with language where no clear coordination on content, by interlocutors, is to be expected. Meaning proper can only be generated and interpreted in accordance with conventions and prevailing standards in the community of reference.

I argue against this form of radical conventionalism and, crucially, against its promise of linguistic interpretation without mindreading, general interpretation and rationalization. I stress how the uncritical following of conventions and standards leads to unacceptable results - namely, the interruption of communication and mutual understanding - in cases where expressions are used in new or non-standard ways. These cases work as a wedge in my argument. There being a permanent possibility of non-standard uses, interpreters must test for the adequacy of the standard interpretation in all cases, in a perennial effort of rationalization. Conventions of meaning, I maintain, are generally insufficient for linguistic communication.

7. Radical Anti-conventionalism: Davidson on Meaning Conventions

There is a tempting view of linguistic practices and interpretation that Davidson identifies and denounces as inadequate in a number of texts from the eighties and nineties. He sums it up in passages like the following:

> … the concept of a language I opposed. It was this: in learning a language, a person acquires the ability to operate in accord with a precise and specifiable set of syntactic and semantic rules; verbal communication depends on speaker and bearer sharing such an ability, and it requires no more than this. (Davidson, 1994b, 110)

Views of the type in question - endorsed in different forms, by different authors, with different proposals - place some notion of *regularity*, *convention* or *rule* at the core of linguistic competence, and reduce interpretation to the exercise of a very specific, and potentially automatic[134], ability - the ability to calculate the standard meanings of uttered expressions in accordance with some system or code commonly known by all participants. Such a straightforward method would release interpreters - at least in some cases - from the need for any additional inquiry into the speaker's mind and background. To argue against this view, Davidson tries to prove that these notions - *convention, regularity* or *rule* - are not, after all, so important or central to successful linguistic interchanges.

In this chapter, I start with a first introduction of Lewis' notion of *convention*, a notion that will anchor much of the discussion in these last two chapters where ample attention will be paid to the social structures that support or facilitate communication and interpretation.

Next, I turn to a second notion, Davidson's *first meaning.* Davidson motivates it first as a key piece in his account of situations where the interchange of literal meanings among speakers is successful despite the fact of words not being used to mean their conventional or

[134] Davidson invokes, for instance, the idea of a "portable interpreting machine set to grind out the meaning of an arbitrary utterance." (Davidson, 1986, 107)

standard meaning. I defend that this notion is also particularly fit to describe what takes place in events of interpretation from scratch and of first language acquisition. Lastly, I reveal how *first meaning* is also useful to clarify the type of meanings that are central to this dissertation. What I defend when I defend that interpretation requires rationalization is that interpreters' investigation of first meanings requires rationalization.

Equipped with these notions, I can finally proceed with the exploration of the crucial question in these last two chapters: how much can the appeal to conventions mitigate the need to engage in rationalization and mindreading in interpretation?

Davidson does not dispute the pervasiveness and practical convenience of conventions of meaning in verbal communication. What he disputes is that there is a connection deeper than that; what he disputes is a certain conception of language and linguistic practices as being intrinsically conventional. In defense of an alternative conception, he sets himself to show that conventions are neither sufficient nor necessary for linguistic communication. I present and inspect several (more or less straightforwardly) Davidsonian arguments to that effect. I conclude that there is a good case for the general insufficiency of conventions in interpretation, but that the arguments for non-necessity fall short of demonstrating a strong and general dispensability of conventions.

7.1. LEWIS' NOTION OF CONVENTION

There is a discernible emphasis on *convention*, even if Davidson often shifts from that notion to others closely related - that is, there is also talk of *regularities*, *knowledge of regularities*, *conventions*, *rules*, *norms*, and more. I will focus exclusively on convention as it includes all or most of what Davidson has in mind in his arguments and positions. Davidson, along with most others in this discussion, refers to David Lewis' analysis of the notion (Lewis, 1969 and 1975).

According to Lewis, conventions arise as rational solutions to coordination problems. Coordination problems are problems involving more than one agent, where there is a predominant coincidence of interest among the participants, and where the success of each agent in reaching some desired outcome is dependent upon her ability to concert her course of action with those of the other participants. Participants must form their strategies based on

their expectations about the others. No individual move can warrant a favorable result in such situations; only coordinated actions can pull it off.

Lewis provides several simple and illuminating examples of coordination problems such as *meeting someone*, *resuming an interrupted phone conversation*, *rowing together*, *driving on two-lane roads* and more (Lewis, 1969, 5 ff.).

> Suppose you and I both want to meet each other. We will meet if and only if we go to the same place. It matters little to either of us where (within limits) he goes if he meets the other there; and it matters little to either of us where he goes if he fails to meet the other there. We must each choose where to go. The best place for me to go is the place where you will go, so I try to figure out where you will go and to go there myself. You do the same. Each chooses according to his expectation of the other's choice. If either succeeds, so does the other; the outcome is one we both desired. (Lewis, 1969, 5)

In the example, we see there are various possible solutions that are favorable to both participants - they can meet at numerous different places. This is an essential feature of the type of situation where a convention - in the Lewisian sense - can play a role. This allows for the arbitrary nature of conventions - a number of concerted strategies would do equally well for all involved, but they have to converge on only one.

Lewis characterizes participation in a convention as a rational solution to problems of this sort - at least to some. A convention might strike one as a disproportioned solution for a *meeting problem* - usually *meeting problems* are not so recurrent and regular as to call for a conventionalized response as people usually settle for mere *ad hoc* arrangements. But consider, in contrast, a further example:

> Suppose several of us are driving on the same winding two-lane roads. It matters little to anyone whether he drives in the left or the right lane, provided the others do likewise. But if some drive in the left lane and some in the right, everyone is in danger of collision. So each must choose whether to drive in the left lane or in the right, according to his expectations about the others: to drive in the left lane if most or all of the others do, to drive in the right lane if most or all of the others do (and to drive where he pleases if the others are more or less equally divided). (ibid., 6)

In this case, the problem is wide and recurrent enough to justify a full-fledged convention. At the root of it all, it must happen that some particular coordination strategy is collectively focused - this can happen in a number of ways, from explicit agreement to some accidental precedent, or natural saliency. The rest follows rationally. Agents that start with

adequate expectations about each other and a common interest, choose their actions in accordance, and manage to coordinate and attain common success. Good results reinforce the initial mutual expectations, new successes grow more likely each time, and participants' behavior becomes more and more regular. At some point, a *convention,* in the Lewisian sense, will be in force. Here is a late version of Lewis' analysis, with some trimming:

> A regularity *R*, in action or in action and belief, is a *convention* in a population *P* if and only if, within P, the following six conditions hold. (Or at least they almost hold. A few exceptions to the "everyone"s can be tolerated.)
>
> (1) Everyone conforms to *R*.
>
> (2) Everyone believes that the others conform to *R*.
>
> (3) This belief that the others conform to *R* gives everyone a good and decisive reason to conform to *R* himself. (...)
>
> (4) There is a general preference for general conformity to *R* rather than slightly-less-than-general conformity. (...)
>
> (5) *R* is not the only possible regularity meeting the last two conditions. There is at least one alternative *R′* such that the belief that the others conformed to *R′* would give everyone a good and decisive practical or epistemic reason to conform to *R′* likewise; (...)
>
> (6) Finally, the various facts listed in conditions (1) to (5) are matters of *common* (or *mutual*) *knowledge*: they are known to everyone, it is known to everyone that they are known to everyone, and so on. (Lewis, 1975, 164-5)

We can see that, for instance, the last example concerning two-lane driving fits very well. Take, for instance, continental Europe. There, every driver drives in the right lane, and everyone expects all other drivers to do the same. The belief that all other drivers will be driving in the right lane gives each driver a good and decisive reason to drive in the right lane. No one would profit from exceptions to this regularity - that is, if most people already drive in the right lane, everyone prefers that all drivers do the same. It would do just as well if all drivers were instead to drive in the left lane. The original choice between left or right is arbitrary as there are no (evident) reasons to prefer one side or the other. Lastly, all this is common knowledge[135, 136] to all involved: all drivers know all the facts stated above, and

[135] Lewis settles for *potential* common knowledge, that is, knowledge "available if one bothered to think hard enough" (Lewis, 1975, 165)

[136] I agree with Daniel Nolan when he notes that condition (2) seems somewhat "redundant, given that all the conditions must be common knowledge in the population P" (Nolan, 2005, 161). It may generate some

know that all drivers know them, and know that all drivers know that all drivers know them, *and so on*. This last condition is meant to *stabilize* the convention. Remember that these are situations where agents must choose their actions based on their expectations about the other agents. Common knowledge of all these facts allows each agent to replicate the others' reasonings and, in general, to grasp the rationality and goodness of the whole arrangement, hence confirming the agent in the persuasion that driving in the right lane is both the thing to be expected and the right thing to do.

Lewis uses his results on the general notion of convention as the groundwork for his own foundational semantics which presents speakers' use of their language as something that is conventionally determined. This corresponds to one way of applying Lewis' notion to affirm a conventional relation of words to their meanings but not the only possible one. Lewis built his account on the notion of truth. According to his proposal, what relates a speaker to her particular language, or idiolect, and her words and sentences to their particular meanings, is a convention of *truthfulness and trust* in that language. A regularity *R*, in action and belief, is a *convention of truthfulness and trust in £* in a population *P* if and only if,

(1) Everyone conforms to *a regularity of truthfulness and trust in £*, i.e. "the members of P frequently speak (or write) sentences of £ to one another. When they do, ordinarily the speaker (or writer) utters one of the sentences he believes to be true in £; and the hearer (or reader) responds by coming to share that belief of the speaker's (unless he already had it), and adjusting his other beliefs accordingly." (Lewis, 1975, 167)

(2) Everyone believes that the others conform to the regularity in question, i.e. "the members of P believe that this regularity of truthfulness and trust in £ prevails among them." (ibid., 167)

(3) This belief that the others conform to the regularity of truthfulness and trust in £ gives everyone a good and decisive reason to conform to it. A practical reason, when in the role of a speaker; an epistemic reason when in the role of a hearer.

(4) "There is a general preference for general conformity to the regularity of truthfulness

---

confusion, but this redundancy is also important to highlight the crucial importance of the first order expectations regarding others' behavior captured by condition (2). The proviso, noted in the previous footnote, regarding the potential character of the common knowledge concerns mainly more complicated and higher order expectations. That first belief or expectation that all participants conform to the regularity in question is more urgent than the others, and to miss it is to be outside the convention - I will develop this point further ahead.

and trust in £. … This general preference is sustained by a common interest in communication.” (ibid., 168)

(5) “The regularity of truthfulness and trust in £ has alternatives.” (ibid.)

(6) “Finally, all these facts are common knowledge in P. Everyone knows them, everyone knows that everyone knows them, and so on. Or at any rate none believes that another doubts them, none believes that another believes that another doubts them, and so on.” (ibid.)

As noted, this is but one way of employing Lewis’ analysis of convention to affirm a conventional relation between words and their meanings. I will not consider Lewis’ reasons behind his adoption of this particular form. I will, however, change to a simpler and more obvious alternative. It would probably not fit all the purposes and considerations Lewis had in mind with his proposal, but it is good enough to frame the discussion ahead while avoiding the cumbersome *truthfulness and trust* idioms, and a potentially polemical approximation between illocutionary and perlocutionary effects. Here is what I have in mind: instead of a *convention of truthfulness and trust in £*, I will be discussing a convention *of using sentences/expressions in accordance to their meaning in £.* The change only concerns the regularity in question. Condition 1, in particular, should now read something like:

(1) Everyone conforms to *a regularity of using sentences/expressions in accordance to their meaning in £*. That is, if someone utters some sentence *s,* that person means by it what *s* means in £; if someone hears it uttered, that person responds by coming to take the speaker to have meant by it what *s* means in £.

Here is a simplified version of the whole thing: a regularity *R*, in action or in action and belief, is a *convention of using sentences/expressions in accordance to their meaning in £* in a population *P* if and only if,

(1) Everyone conforms to a regularity of using sentences/expressions in accordance to their meaning in £.

(2) Everyone believes that the others conform to the regularity of using sentences/expressions in accordance to their meaning in £.

(3) This belief that the others conform to the regularity of using sentences/expressions in accordance to their meaning in £ gives everyone a good and decisive reason to conform to it.

(4) There is a general preference for general conformity to the regularity of using sentences/expressions in accordance to their meaning in £.

(5) The regularity of using sentences/expressions in accordance to their meaning in £ has alternatives.

(6) All these facts are common knowledge in P.

These conventions relating populations to their languages comprehend a number of subsidiary conventions[137], in particular, an infinite number of conventions governing the use of sentences of £ - after all, the convention is realized only by the particular uses of such sentences. Take, for instance, "Snow is white". Following the previous line, we should say that a regularity *R*, in action or in action and belief, is a convention, in a population P, of *using "Snow is white" to mean that snow is white* if and only if,

(1) The members of P conform to *the regularity of taking "Snow is white" to mean that snow is white*. That is, if someone utters "Snow is white" that person means that snow is white; if someone hears it uttered, that hearer responds by taking the speaker to have meant that snow is white.

(2) Everyone believes that the others conform to *the regularity of taking "Snow is white" to mean that snow is white.*

(3) This belief that the others conform to *the regularity of taking "Snow is white" to mean that snow is white* gives everyone a good and decisive reason to conform to it.

(4) There is a general preference for general conformity to *the regularity of taking "Snow is white" to mean that snow is white*.

(5) *The regularity of taking "Snow is white" to mean that snow is white* has alternatives.

(6) All these facts are common knowledge in P.

With this notion of convention in place, we are now in a better position to consider Davidson's claims about its dispensability and insufficiency for linguistic communication. First, however, we must clarify the ideas of linguistic communication and interpretative success at stake in his arguments and claims, as well as the notion of meaning involved.

[137] We will need to get to these smaller components to be able to apply these ideas in the discussion ahead which will be rooted in occasional discrepancies among speakers.

7.2. FIRST MEANING

Davidson coined the expression "*first meaning"* (Davidson, 1986, 91) to delimit the notion of meaning central to all this discussion - a notion whose theoretical importance, even indispensability, in various relevant contexts is often underappreciated.

Davidson is clear about his conviction that the crucial thing in meaning and communication is being understood as one intends to be understood. In his words, the "intention to be taken to mean what one wants to be taken to mean is ... the only aim that is common to all verbal behavior" (Davidson, 1994b, 120)[138]. All successful instances of communication involve the satisfaction of some communicative intention and all conveyed (non-natural) meaning must be specified there, in its correspondent intention.

Even so, the same utterance can be intended to mean various things. Besides what the speaker intends her words to be taken to mean, there will often be additional meaning effects intended by the speaker - take, for instance, cases of irony or metaphor. Davidson wants to preserve this distinction between literal meaning and speaker meaning but he does not accept to have it built over an identification of literal meanings and conventional or standard meanings.

> We want a deeper notion of what words, when spoken in context, mean; and like the shallow notion of correct usage, we want the deep concept to distinguish between what a speaker, on a given occasion, means, and what his words mean. The widespread existence of malapropisms and their kin threatens the distinction, since here the intended meaning seems to take over from the standard meaning. (Davidson, 1986, 91)

Davidson's reflection on the role of conventions in linguistic communication is intensely focused on the occurrence of *misuses* of linguistic expressions - in particular, malapropisms - that still do not disrupt communication. Speakers often use words intending them to mean something different from their *standard meanings*[139] and "get away with it"

---

[138] See also Davidson, 1993, 171, and 1986, 92-3 and 98-9.

[139] Throughout this chapter I will be sticking to this formulation - "standard meaning" - to refer to the usual meaning of a linguistic expression in some community of reference. Authors often use other terms, such as "conventional meaning" or even "dictionary meaning", to the same effect. I especially avoid "conventional meaning" in this *standard meaning* sense to mark the difference and prevent confusions with the *Lewisian notion of conventional meaning* characterized above. The two notions are importantly divergent, and I will stress this distinction by often articulating both adjectives - "standard" and "conventional" - instead of just one

(Davidson, 1986, 98), i.e. are understood by their audiences as meaning what they actually intended to mean. Mrs Malaprop produces an utterance of "a nice derangement of epitaphs" and is successfully interpreted as meaning what she intends to mean, i.e. *a nice arrangement of epithets* (ibid., 103-4). Archie Bunker uses "monogamy" and is successfully interpreted as meaning what he intended to mean, i.e. *monotony* (ibid., 90). Goodman Ace (ibid., 89) explores similar replacements - using, for instance, "granite" instead of "granted" and "baffle" instead of "battle" - only that he does so intentionally and much more densely.

With *first meaning,* Davidson is forging a notion that still allows for the distinction between *literal meaning* and *speaker meaning* while accommodating the fact the former might not be standard or conventional - *epithets*, *monotony*, *granted* and *battle* are meant literally in the cases above. *First meaning* corresponds to a notion of *intended, potentially non-conventional and non-standard, literal meaning of an utterance*. Davidson wants, in particular, to make sense of cases where speakers get their words understood in non-conventional and non-standard ways and still manage to add to that some extra layer of meaning. Lepore and Ludwig elaborate an example:

> Hearing Mrs Malaprop, we may reuse some of her misused words in fun, saying, "And that's a nice derangement of words," intending them to be understood as meaning what they were misunderstood to mean (and then meaning them ironically)... Our utterance in the imagined context will be understood to mean "And that's a nice arrangement of words," but only ironically, that is, we will be taken to have meant, by so meaning with these words, that it was not a nice arrangement of words. (Lepore & Ludwig, 2005, 265-6)

The first meaning of an utterance will often correspond to the expression's standard meaning. Even more often, the first meaning will "come first in the order of interpretation" (Davidson, 1986, 91-2)[140]. Ultimately, however, it is the speaker's intention that specifies it. First meanings are determined by what the speaker intends his words to mean, i.e. by what Davidson refers to as the speaker's "semantic intention" (Davidson, 1993, 170-1).[141]

---

as if they were synonymous. I will further elaborate on this divergence between conventional and standard meanings in the next chapter.

[140] Some utterances wear their meanings on their sleeves but frequently hearers are required to search beyond the literal meaning of the words for additional meanings. In both cases hearers tend to start their interpretation from the words' first meanings. In the latter cases the grasp of first meaning will often constitute a first indispensable step in trying to infer or otherwise discover the rest.

[141] In 1986, in "A Nice Derangement of Epitaphs", Davidson tells a slightly complicated story. He appeals to the order of the speaker's intentions behind the utterance. Building on the ideas that "[t]he intentions with which an

First meaning is *linguistic meaning* (Davidson, 1986, 93), in the sense that it is a form of *non-natural meaning* (Grice, 1957) that is then specifically differentiated by being articulated by means of expressions integral to a complex system, a language, that allows for the properties of compositionality, systematicity and creativity characterized in section 1.1. of this dissertation. First meanings appear thus at a crucial intersection between mind and language. They first link interpersonal understanding to a verbal code, dramatically improving the first and infusing the second with actual content.

In a very important sense, first meanings are prior and necessary to any form of stabilization of expression meaning, be it by means of conventions or, simply, as community standards. In the normal cases[142] the meaning convention will have emerged naturally from antecedent cases of successful interchange of the correspondent first meaning. The convention's work is not to invent the solution for the coordination problem; it is to ensure or promote its recurrence. As explained above, a condition for any convention is that some particular coordination strategy is already collectively focused by the participants - this introduces the type of conduct that is to become regular with the convention. Conventions can only be built upon such antecedents and, of course, meaning conventions are no exception. The emerging of any meaning convention is dependent upon some number of prior occurrences where participants must have managed to coordinate on taking the expressions in question to mean the relevant particular meaning. In those prior occurrences

---

act is performed are usually unambiguously ordered by the relation of means to ends" (1986, 92), Davidson explains that first meanings are the ones specified by the first intention of the right sort in the sequence. To illustrate, Davidson recalls Diogenes' story with Alexander: "Suppose Diogenes utters the words 'I would have you stand from between me and the sun' (or their Greek equivalent) with the intention of uttering words that will be interpreted by Alexander as true if and only if Diogenes would have him stand from between Diogenes and the sun, and this with the intention of getting Alexander to move from between him and the sun, and this with the intention of leaving a good anecdote to posterity" (ibid). To distinguish the intentions that matter, he then invokes a distinctive *feature* of meaning intentions, according to Grice's analysis. Davidson explains that some of the listed intentions will be "Gricean intentions to achieve certain … ends through Alexander's recognition of some of the intentions involved" (ibid.). When Diogenes utters the words "I would have you stand from between me and the sun", he does so with "the intention of uttering words that will be interpreted by Alexander as true if and only if Diogenes would have him stand from between Diogenes and the sun" by means of Alexander's recognition of this intention of Diogenes. When Diogenes produces his utterance he does so with "the intention of getting Alexander to move" again by means of Alexander's recognition of this further intention of Diogenes. These are the intentions that matter. Then, the "first intention in the sequence to require this feature [i.e. the first meaning intention in the sequence] specifies the first meaning" (ibid.).

In his 1993 *Locating Literary Language*, he states it more simply. He explains that the relevant intention is "the first intention that has to do with what words mean, or are intended to mean, is the intention to speak words that will be assigned a certain meaning by an interpreter." (Davidson, 1993, 172-3)

[142] The only exceptions will be cases where the relevant regularity is agreed upon without being actually exercised. Speakers can talk among themselves to coin a word and decide how they will use it before actually using it in that way. These cases will be very exceptional.

the audience must already grasp that the speaker intends her words to mean what she intends them to be taken to mean. In other words, in those prior occurrences what is achieved is already the successful interchange of first meanings. Conventional meanings develop out of first meanings. With this respect, it is common to understand Lewis' work on meaning conventions as a study on the processes through which speaker meaning can evolve into expression meaning - see, in particular, Avramides (1997) and Schiffer (2017).

The need for a notion of first meaning can be further confirmed when attention is paid to the situations of radical interpretation and of first language word learning. In both types of situations it is the utterances' first meanings that are the primarily important type of expression meaning and what hearers and interpreters strive to find out. We are well placed to understand this since radical interpretation and word learning are precisely what we have been paying attention to in all the six previous chapters.

The child's or radical interpreter's path to any type of expression meaning - *conventional* meaning, *standard* meaning, *dictionary* meaning - must start with the discovery of first meanings. At the beginning, there is no question of conventional meaning in a Lewisian proper sense since child and interpreter share no convention with the speaker. Furthermore, in such an initial and fundamental moment it makes no difference to child or interpreter whether the speaker's intended association between expression and meaning corresponds to a widespread practice in the community, or figures in a dictionary. Child and interpreter could not possibly tell one case from the other. Accordingly, what we need here is a notion of meaning that is prior and neutral with respect to standard uses, dictionaries and conventions.

Child and interpreter must inquire after the speaker's communicative and semantic intentions. The speaker's use of her words is the sole clue available to their meaning but only, as we saw, inasmuch as her communicative intentions - and other implied thoughts - behind such uses are inferable. Hence, to account for these types of situations we need a notion of meaning that is determined by the speaker's intentions. Furthermore, at the beginning, interpretative successes will be mostly confined to the simplest uses of words. Speakers will tend to try to convey their meanings in the plainest and straightest way at their disposal. For instance, they will avoid confusing the novice with attempts at conversational implicatures. Even if they did not, however, child and interpreter would likely miss their points. What we need to account for these types of situations is a notion of meaning that is determined by the

speaker's communicative intentions of the most elementary linguistic type, a notion like *first meaning.*

It is opportune here to return to the question of foundational semantics to address an apparent tension between the idea, put forward in this section, that first meanings are determined by what the speaker intends his words to mean, as long as the interpreter manages to grasp that intention, and Davidson's position on the constitutive radical interpretability of meanings and thoughts. This illusion of a threat is easily dispersed. As defended above, the claim that meanings are determined by their radical interpretability is compatible with the existence of concurrent factors of determination. In Chapter 3, I maintained that Davidson's view is richer and more comprehensive than the kind of truncated *Interpretivism* with which he is often associated. Davidson argues for a concert of perspectives – first- and third-personal - over thoughts and meanings. It is this idea of concert that allows us to make sense of diverse but convergent determining or constitutive factors.

Lastly, it is important to note that this notion of first meaning subsumes the type of meanings that are central to this dissertation. I explained in the introduction that I am confining my investigation into interpretation to the most basic cases of linguistic interpretation. I take these cases to be those where the interpreter manages to recover the literal meaning of the words of some utterance, when this meaning happens also to be the speaker's ultimately intended meaning.

I am interested in linguistic practices and meaning production and interpretation as processes instrumental in, and subordinated to, communication and understanding. I follow Davidson when he claims[143], almost trivially or vacuously, that there is verbal understanding and communication when interpreters' recognize agents to have intended to mean what they have intended to mean. First meanings introduce the linguistic element into the affair as the relevant communicative intentions must be intentions for the utterances produced to be recognized as specific linguistic forms carrying specific linguistic meanings but, importantly, they do so without committing us to the superfluous burden of shared languages, standards or dictionaries. Furthermore, first meaning is basic to all verbal communication. In this way, the notion allow us to associate, in a very fundamental way, all sorts of successful occasions of linguistic interchange, from the most common and regular verbal transaction among mature

[143] See, for instance, Davidson, 1994b, 120.

competent speakers of the same natural language, to the successes of a radical interpreter in understanding some remote native, while not forgetting the cracking of some occasional twist of words, intentional or not. What all these cases have in common is that the words articulated are recognized by the audience as meaning what the speaker intended them to mean.

I am stopping at this basic level. As noted already in the introduction, in this study I am leaving aside the consideration of more sophisticated pragmatic effects - which, anyway, are almost consensually taken to engage the audience in some form of rational calculation - to focus exclusively on that first level of meaning that speakers intend to bestow upon their words.

7.2.1. FIRST MEANING, UTTERER'S MEANING, UTTERANCE-TYPE MEANING AND WHAT IS SAID

There are evident affinities, but also differences, worth pointing out between Davidson's *first meaning* and some of the notions of meaning put forward by Grice. This is opportune because it allows us a more complete and integrated picture of Davidson's notion and worries, and of their place in a wider context. It will also help us, once again, to delimit the type of meanings that are the object of this dissertation.

As noted, *first meaning* is presented by Davidson as a species of *non-natural meaning* (Davidson, 1986, 93), and very close to Grice's fundamental notion of *utterer's meaning* or *speaker meaning*. In Grice's conception all forms of non-natural meaning are derivative from those semantically foundational events in which an agent means something by doing something (producing some utterance) with a complex intention of a certain type[144]. Davidson's first meanings share an extremely important feature with utterer's meanings, the centrality of communicative intentions and of their recognition in the determination of those meanings.

Unlike Grice, however, Davidson holds no hope for an apt reduction of any type of non-natural meaning to non-semantic notions[145]. We find that Davidson's minimal

[144] See Grice, 1957, 217 and 220, and Grice 1968, 117.
[145] See Davidson, 1994, 121, fn.13

conception of a communicative intention as an "intention to be taken to mean what one wants to be taken to mean" (Davidson, 1994, 120) fits well enough Grice's general and rough definition of "*A* meant something by *x*" as "*A* intended the utterance of *x* to produce some effect in an audience by means of the recognition of this intention" (Grice, 1957, 220). The disagreement starts only afterwards, with Davidson's reluctance towards the idea developed by Grice that the intended effects can be specified without the use of any notion of meaning - in terms, in particular, of beliefs or actions[146].

First meanings and utterer's meanings part ways when interlocutors go beyond and against the literal meanings of the expressions they use. While *utterer's meaning* is general enough to accommodate all that the utterer intends to communicate with her utterance, first meanings are confined to the intended literal meaning of her words.

First meaning, as well*,* has an extended life beyond these cases of coincidence of intentional and literal meaning. As explained above, Davidson has work for it to do also in those situations involving supplementary layers of signification. First meanings are still present and instrumental in supporting interlocutors reaching further, non-literal, meaning effects, such as in cases of irony, hints, or metaphors. This brings to the fore a second partial affinity now with a different type of meaning notion. From this angle *first meaning* appears in line with Grice's notion of *expression meaning* (Grice, 1968) - or *timeless meaning*, or *linguistic meaning*, to indicate just two alternative labels Grice uses to refer to the same thing. Both Davidson's *first meaning* and Grice's *expression meaning* make room for the fact that the meaning of expressions preserves some degree of autonomy from their uses and their user's ultimate communicative intentions. Grice, however, does not touch the issue of *timeless meaning* needing to be *bestowed* or *confirmed* onto the words uttered by means of a speaker's intention to that effect, as is the case with Davidson's first meanings.

It is also important to note that, despite conceding on the habitual preponderance of conventions and standards in the fixing of expressions' meanings[147], Grice explicitly denies,

[146] See Grice, 1968, 123, and Grice, 1969.
[147] Here, for instance, is one passage illustrating it: "In the sense in which I am using the word say, I intend what someone has said to be closely related to the conventional meaning of the words (the sentence) he has uttered. Suppose someone to have uttered the sentence He is in the grip of a vice. Given a knowledge of the English language, but no knowledge of the circumstances of the utterance, one would know something about what the speaker had said, on the assumption that he was speaking standard English, and speaking literally." (Logic and Conversation, 25)

in a later text, that this frequent association should be understood as a form of essential dependency.

> I do not think that meaning is essentially connected with convention. What it is essentially connected with is *some way of fixing what sentences mean*: convention is indeed one of these ways, but it is not the only one. (Grice, 1982, 298; the emphasis is mine)

In placing convention merely on a par with other forms of fixing linguistic meanings, Grice appears, one more time, interestingly close to Davidson.

Lastly, it is also useful to compare the notion of *first meaning* with the notion of *what is said*. This last notion is designed by Grice to capture the type of meanings at the intersection of *expression meaning* and *utterer's meaning*[148]. On those occasions where first meanings are the only meanings speakers intend their utterances to convey, first meanings seem to perfectly match *what is said*. As it happens, it is precisely this subset of first meanings that correspond to *what is said* that is central to the most basic type of linguistic communication, the type that this dissertation is especially concerned with.

7.3. DAVIDSON'S ANTI-CONVENTIONALISM

Davidson's position on meaning conventions is complex as he expresses different forms of rejection to different conceptions of the alleged conventional nature of linguistic practices.

First of all, he never rejects that there actually are meaning conventions, nor that they are pervasive and practically quite useful in real instances of communication (cf. Davidson, 1994b, 110; and 1982a, 278). What Davidson rejects is, on the one hand, the *necessity* of such conventions for communicative purposes and, on the other, its *sufficiency* (cf. Davidson, 1994b, 110).

Second, we must also pay attention to the different types of conventions he refers to. We must distinguish at least two types of regularities considered by Davidson. The first type of regularity considered determines a strong form of sharing. It consists of "speaker and hearer mean[ing] the same thing by uttering the same sentences" (Davidson, 1982a, 276).

---

[148] Grice, 1969, 87-8 and Grice, 1968, 120-1; See also Neale, 1992, 520-1 and 554-6.

Davidson defends that “such conformity, while perhaps fairly common, is not necessary to communication” (ibid.). He immediately points to the obvious counterexamples: “[e]ach speaker may speak his different language, and this will not hinder communication as long as each hearer understands the one who speaks” (ibid.). Examples include cases of speakers of officially different languages and cases of speakers that consistently pick different words and formulations within the same official language or dialect.

To avoid the kind of counterexamples just noted, Davidson moves his focus to a second type of regularity. It determines a much weaker form of sharing and consists of “speaker and hearer … assign[ing] the same meaning to the speaker’s words” (ibid., 277). This time, it does not matter which expressions each interlocutor picks in her utterances, as long as both speaker and hearer are ready to interpret them in the same way. This is the type of regularity we will be focusing on.

A further idea is important, as Davidson notes: “Regularity in this context must mean regularity over time, not mere agreement at a moment” (ibid., 278). Also, whether or not the regularities in question correspond to any official language, or constitute some standard in a wide enough community, is not a particularly discernible worry in Davidson’s approach. He centers his argumentation on the general case, conventional meanings, of which *official language conventions* and *large community language conventions* would constitute the most typical examples.

Davidson states his anti-conventionalism in yet another form. In his 1986 *A Nice Derangement of Epitaphs,* he famously declares that “there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed” (Davidson, 1986, 107). Naturally, Davidson has no intention of refusing the existence of language, or even languages, *tout court*. What he is refusing is only a certain conception of it, even if one that enjoyed ample acceptance, explicit or tacit, from theorists - we can call it “*the stiff conception of language*” to facilitate further referencing. Davidson characterizes it invoking three features. This purportedly flawed conception of language takes the meanings involved in instances of successful verbal communication to be *systematic*, *shared* and *governed by learned conventions or regularities* (ibid., 93). Davidson defends that we should let go of the third feature.

Meanings, in his revised conception, must still be systematic and shared. Only a rich and articulated enough system can offer the expressiveness characteristic of verbal

communication. Furthermore, speaker and interpreter still need to converge on the first meanings of the words used during their interchange if communication is to count as effective. What Davidson questions is the need for such a convergence to be secured beforehand by means of both interlocutors bringing to the conversation an acquired, common and standing knowledge of the conventions and regularities of the language. Davidson is convinced that the frequent cases of successful communication undisturbed by semantic innovations - which he illustrates with his examples of malapropisms - force us to reconsider and ultimately eliminate the requirement.

Davidson formulates his revision of the stiff conception with the help of the notions of *prior theory* and *passing theory* (ibid., 100-1). The theories in question would be descriptive meaning theories, truth theories if we follow Davidson's program. Prior theories model the competence or dispositions to linguistic behavior that each participant brings to the conversation. The interpreter's prior theory should characterize his dispositions to interpret his interlocutor prior to actual conversation, and the speaker's prior theory should characterize how she expects the interpreter to understand her words. Prior theories, Davidson defends, may be discrepant without any loss of communication. Passing theories model verbal dispositions during the interchange, as the relevant news is gathered and processed. The interpreter's passing theory characterizes how he actually ends up interpreting the speaker's words, while the speaker's passing theory characterizes how she intends him to interpret them. It is passing theories that need to coincide if communication is to be successful. And that is enough to ensure that speakers' words are taken to mean what they intend them to be taken to mean.

According to Davidson's revised view, a language is at most a fleeting thing, fixed only for the instant of communication, and not a stable system, shared among all speakers and accounting for their linguistic competence, something to be "learned, mastered or born with" (ibid., 107). What speakers carry with them from situation to situation, allowing them to reach a good enough coordination each time, cannot be identified with either prior (ibid., 103) or passing theories (ibid., 102). It is not, in fact, something that could fairly be described as knowledge of a language, whether one imagines it encapsulated in a module or the cognitive possession of the whole mind and person. Even if rooted and dependent on a general and structural competence comprising "a basic framework of categories and rules, a sense of the way English (or any) grammars may be constructed, plus a skeleton list of

interpreted words for fitting into the basic framework" (ibid., 104), speakers' performances reveal more than the simple employment of any form of standing knowledge. They reveal the continuous exercise of a dynamic, complex, general and high-order capacity, the speaker's general intelligence, able to serve them in communication as well as in other forms of theory building and mundane navigation (ibid.107). In other words, and bringing it closer to home, they reveal a continuous exercise of rationality and rationalization.

This picture constitutes an enlarged rendition of the same anti-conventionalist position described above. In allowing prior theories to diverge, Davidson sustains the insufficiency and unnecessariness of conventions and previously learned regularities for communication. The conventions that are in place, where speaker and interpreter's prior theories intersect, are not enough to ensure communication. Prior theories must be revised or confirmed into passing theories. Speakers achieve this not with the help of further conventions and strict strategies but by making free and ingenious use of their broad rationality in exploring varying and unchartable provisions of miscellaneous knowledge.

It is time now to turn to Davidson's arguments in defense of this view.

7.3.1. ARGUING AGAINST CONVENTIONALISM

In support of his revised view of language and verbal communication and of his claims that conventions are neither necessary nor sufficient for linguistic communication, Davidson invokes the fact that, sometimes, speakers must figure out the meaning of their interlocutors' words without relying on previous knowledge of any prevailing convention. Speakers can do it, and often have no option but to try it (Davidson, 1982a, 278; 1986, 89-90; 1994b, 115 ff.).

Malapropisms are taken as central examples in this argument. The examples elaborated above (section 7.2.) already go to show that, at least occasionally, speakers manage to communicate and understand each other by exploring other means besides shared knowledge of the relevant meaning conventions. There being no linguistic convention of using "baffle" to mean *battle,* knowledge of linguistic conventions is flagrantly *insufficient* to explain how it can happen that on certain particular occasions speaker and hearer manage to converge in this interpretation. From the opposite angle, even without the knowledge of the convention of using "arrangement" to mean *arrangement*, as presumably was the case with

Mrs Malaprop, a speaker can still be understood by her interpreter to mean it - thus showing how conventions can also be *unnecessary.* The important question now is how far these conclusions can be extrapolated. Is there a valid inference from occasional to any general form of insufficiency and dispensability of meaning conventions in communication?

7.3.2. ARGUING AGAINST SUFFICIENCY

Are conventions and knowledge of conventions enough to explain and allow for linguistic communication in its familiar form? If there is successful verbal communication involving the employment of expressions in non-conventional and non-standard ways, then it follows that, at least for these cases, conventions, standards and the correspondent knowledge are insufficient for communication. This is the argument that Davidson has in mind against sufficiency and, accordingly, he illustrates the existence of such cases of non-conventional successful communication with the examples provided above. Davidson himself does not offer any other clear argument on insufficiency; he actually appears much more interested in the second aspect of the problem - necessity.

As things stand, the insufficiency of conventions and standards is secured only at a very local and occasional level. Some extraordinary occasions require speakers to coordinate beyond what is previously fixed and shared among them. But it would seem that, outside those rare occasions, learned conventions and standards still suffice for communication. We can easily agree that the occasions might not be so rare after all, and that they encompass more than malapropistic uses. There are many other cases of improvisation and innovation: new names, new words, new idioms, new uses. Nonetheless, if insufficiency and the opportunity or need to invest more in communication than one's knowledge of fixed standards and conventions were keyed only to such special moments of innovation, they would still be confined and of little expression in the whole picture of language use.

Pietroski, for one, in a commentary sympathetic to Davidson's own position, expands on the topic and reveals an argument supporting the generalization of insufficiency. It appears contained in a single paragraph that is worth quoting at length.

> Once the distinction between prior and passing theories has been drawn, I take it that the conceptual distinction matters even if the deliverances of prior and passing theories are the same. Not modifying a prior theory is, on my reading of Davidson, just as much an interpretive decision as

> making a modification; though in the former case, the 'decision' will typically not be associated with any conscious processing or feelings of 'dissonance.' Successful communication is always a matter of converging *passing* theories; and general intelligence is always implicated here, if only by giving 'tacit approval' to the deliverances of prior theories. Moreover, even if such tacit approval is often granted, one cannot speak of rules *governing* passing theories. For the 'rules' could always be overridden in cases of the Malaprop/Donnellan sort; and one cannot capture the extent of these cases formally or in advance. (Pietroski, 1994, 105)

Pietroski points out that the simple drawing of the conceptual distinction between prior and passing theories creates room for the possibility of a permanent actualization of linguistic dispositions taking place when speakers engage in verbal communication, whether or not changes are actually consummated in the process. He claims that an “interpretive decision”, explicit or tacit, is in any case called for, be it to confirm or modify the prior theory, and that this can be so even if the interpreter is unaware of it, phenomenologically speaking. Lastly, he stresses the fact that *the last word* does not belong to the conventions or rules of language, which again supports the idea that there is always interpretation, and central and personal control of the interpreter over how to interpret any utterance.

Lepore and Ludwig find the following argument in their interpretation of Pietroski’s passage: since “the possibility a speaker has not used his words in conformity with public norms is ever-present” (Lepore & Ludwig, 2005, 270), interpreters are called to justify their option in either way, that is, whether they decide to go with the conventional interpretation or not. “But since [this] justification will invoke more than knowledge of conventions for the use of words, and even that the speaker is a member of the appropriate linguistic community, it follows that knowledge of conventional meaning is never sufficient for interpretive success” (ibid., 271).

They find some truth and cogency in this line of reasoning but, at the same time, they do not fail to recognize that “it also seems clear that we routinely and successfully interpret others on the basis of taking them to mean what their words mean according to public norms” (ibid.). They conclude - pertinently, to my mind - that, ultimately, it all “boils down to what we intend by saying that knowledge of conventional meanings is sufficient for interpretive success” (ibid). If what is meant is a) “that sometimes, even often, we are not called upon to revise our view that the speaker speaks with the majority” (ibid.) then the argument does not secure generalized insufficiency of conventions. If, on the other hand, what is meant is b) “that knowledge of conventional meanings all by itself sometimes suffices for interpreting

another as speaking in accord with public norms" (ibid.) then we can take the argument to show that conventions are indeed always insufficient for communication.

I agree with Lepore and Ludwig's assessment of the argument, even if I have some doubts about its actually being contained in this precise form in the passage from Pietroski quoted above. Moreover, they frame the question in terms of *justification*, and I believe we can do better in different and simpler terms. It is not only that knowledge of conventions is insufficient to sustain a conventional interpretation of an utterance *in a justified way*, it is more simple and serious than that. It is that knowledge of conventions is insufficient to sustain a conventional interpretation of an utterance, even without the qualification. As Lepore and Ludwig articulate the *justification* question they make it sound as if justifying or not one's interpretive decision was somewhat optional. They even explain, in a footnote, that the argument is conditional on a specific epistemological position that liberally affirms the need to justify every interpretation. They point out that this view can be challenged: "it might be maintained that our beliefs about what others in our community mean by their words and actions are justified by default: unless circumstances depart from the norm, in some way that we should notice, the beliefs we have automatically are justified without appeal to anything" (ibid., 271, fn.220) In this contrary conception, only sporadically would justifications be called for, and ascriptions would need to "be actively justified only when circumstances depart in certain specific ways from the norm" (ibid.).

I believe we can form a tighter argument for the general insufficiency of conventions, and one that is not dependent on disputable epistemological stances, by letting go of the idea of justification. It stems from the same crucial observation that there is an ever-present possibility that the "speaker has not used his words in conformity with public norms" (ibid., 270). Sometimes, interpreters detect that this is indeed the case. The occasional detection of non-conventional uses already requires some prior form of attention to that possibility. That is, it cannot be that the detection of deviation triggers deeper interpretation because deeper interpretation is already needed for detecting deviation. The interpreter would not detect any departure from the conventional course if he were not both aware of that possibility of deviant uses and actively, however tacitly, on the look out for them. He would simply continue ascribing standard meanings, and suffer, unwittingly, the consequent losses in understanding. This is not what happens, interpreters do detect anomalies and reinterpret in accordance. Furthermore, I must stress that *every* utterance is a chance for deviation. Because

of that, interpreters are required to be continuously on the lookout. Virtually never are they allowed to rest exclusively on their knowledge of the community's norms and conventions. What emerges here is the generalized insufficiency of conventions and the correlative need for interpreters to engage in a continuous exercise of rationality and rationalization.

Lepore and Ludwig's conclusion still holds. Since meaning ascriptions will, very often, conform to the public norm and demand no revision, there is a sense - see reading a) above - in which it is possible to continue holding that conventions are often all that we need. However, since my fundamental point is whether or not rationalization is needed for interpretation, I am only interested in reading b). The question that matters to this dissertation is whether conventions of meaning can be taken to allow interpreters to completely dispense with rationalization at least sometimes. What the last argument appears to show is that on no occasion is the need to rationalize perfectly supplanted by mere appeal to conventions and knowledge of public norms. I will return to this topic of sufficiency in the next and last chapter of this dissertation. For now, it is time to consider Davidson's arguments against necessity.

7.3.3. ARGUING AGAINST NECESSITY

Can speakers do without conventions in their linguistic practices? Davidson's examples of malapropisms clearly show that sometimes speakers manage to understand each other by means other than the exploration of meaning conventions. Once more, the important question that follows is how far can we extend this conclusion? Is this merely an exceptional phenomenon, or is it possible to engage in verbal communication while completely dispensing with the use of conventions?

The examples explored hint at the weaker conclusion. Not only are those non-conventional uses confined to special occurrences and not very widespread in conversation, but they even seem to be possible only when operating against a background of standing linguistic conventions. First, in about all cases, even if to varying degrees, the figuring out of the intended first meaning seems only possible thanks to the prevalence of standards in the linguistic context of the misused expression, be it simply the rest of the sentence or, more than that, some stretch of the ongoing conversation. How would the interpreter discern *monotony* in "monogamy" were it not for the regular interpretation of "We need a few laughs

to break up the …" (Davidson, 1986, 90). Then, there is often also some form of proximity or association either between the expression actually used and the expression standardly used to mean the intended meaning - typically a suggestive sound resemblance as with "epitaphs" and "epithets" - or even between the intended meaning and the meaning standardly associated with the expression employed - as when Davidson conceives the possibility of successfully using "Water!" to mean "Fire!" (ibid., 89).

The examples offered do not support the general non-necessity of conventions, but only a very local form of it. However, in this case - more than with *sufficiency* - Davidson frequently writes in a way that seems to make evident his intention of holding and defending their complete dispensability. Already in 1982, he states his question in the most general terms:

> The question is delicate because it concerns not the truth of the claim that speech is convention-bound, but the importance and role of convention in speech. The issue may be put counterfactually: could there be communication by language without convention? (Davidson, 1982a, 265)

To which, of course, he answers positively, and still with no hint that he means to confine his conclusions to mere local exceptions:

> In conclusion, then, I want to urge that linguistic communication does not require, though it very often makes use of, rule-governed repetition; and in that case, convention does not help explain what is basic to linguistic communication, though it may describe a usual, though contingent, feature. (ibid., 279-80)

We find exactly the same tone in all other papers on the topic - for instance, twelve years later:

> The theoretical possibility of communication without shared practices remains philosophically important because it shows that such sharing cannot be an essential constituent in meaning and communication. (1994b, 119)

In the rest of this section I will follow two different, but closely related, lines of thought, arguably discernible in Davidson, purporting to reach beyond the examples and sustain the stronger claim that conventions are unnecessary in its stronger and general form. I conclude that none manages to secure the point. I dub the first "the argument from radical interpretability" and the second "the argument from the essence of communication".

7.3.3.1. THE ARGUMENT FROM RADICAL INTERPRETABILITY

In making his point against the necessity of conventions, Davidson invokes a distinction between ideal conditions and assertions of principle, on the one hand, and matters of fact and practical considerations, on the other. He uses it to characterize the role of conventions and knowledge of conventions in linguistic practices, affirming that, instead of essential to interpretation and communication, they are merely convenient, frequently employed but ultimately extraneous to the task.

> Knowledge of the conventions of language is . . . a practical crutch to interpretation, a crutch we cannot in practice afford to do without—but a crutch which, under optimum conditions for communication, we can in the end throw away, and could in theory have done without from the start. (Davidson, 1982a, 279)

However, the applicability of this sort of distinction to the discussion at hand is much less clear than might appear at first. When the topic is linguistic communication among humans, the finite nature of the participants is of central importance, and not an accessory problem. Practical considerations are of essence when the question is precisely what is necessary for two limited creatures to reach verbal understanding. There might be some room to negotiate which constraints to reckon and which to leave aside, what is really indispensable and what is simply hard to do without, but any particular conclusion would require detailed discussion and justification, not just a vague reference to "optimum conditions". What would be *optimum*? To have a lot of time and resources? To be a very smart and lucky interpreter? Surely omniscience or even telepathy are out of the question… I do not think Davidson manages to put forward any realistic scenario in which meaning conventions could fairly be described as mere *practical crutches,* where speakers could actually go about communicating without relying on extensive knowledge of regularities in each others' use of language. I will argue that the prime candidate for such a scenario, the situation of radical interpretation, reveals itself not as *convention free* but rather as the very enactment of a process of initiation of the interpreter into a series of conventions of language that he will end up sharing with the speaker.

That radical interpretation, or at least interpretation with less than perfect prior knowledge of our interlocutor's language, is what Davidson has in mind in passages like the

one above, is clear in the context - in particular, for instance, when he justifies the practical importance of "common conditioning" (ibid., 278) with the fact that "we do not have the time, patience, or opportunity to evolve a new theory of interpretation for each speaker" (ibid.). The unarticulated message in these paragraphs is that, at the same time, there are natural and flagrant limits to the possibility of simply resting on acquired knowledge and homology. Accordingly, one might be tempted to conclude that conventions and knowledge of conventions should be unnecessary from the fact that no speaker is born with this knowledge and yet is able to arrive, sooner or later, at the verbal understanding of others. The flaw in this inference is that it misses the fact that the process of arriving at such an understanding consists precisely in the learning or establishing of linguistic conventions. The point is analogous to the one Lepore and Ludwig make with the following observation:

> It might be objected that, in fact, field linguists can break into alien languages. Of course, this is correct. But they do so by figuring out the regularities in the uses of words by their subjects, which is a matter of learning which conventions govern their words in their linguistic community. (Lepore & Ludwig, 2005, 279-80)

Glüer also acknowledges that "a speaker who does not use his words with a certain regularity would not seem to be radically interpretable" (Glüer, 2011, 110; Glüer, 2013, 353)[149]. In the early beginnings of interpretation there is no alternative but for the speaker to use her words in a regular way. In Davidson's words, "[t]he best the speaker can do is to be interpretable, that is, to use a finite supply of distinguishable sounds applied consistently to objects and situations he believes are apparent to his hearer" (Davidson, 1984b, 13). As for the interpreter, he as well must build and test his hypotheses upon the assumption that the

[149] Glüer (2011, 111; 2013, 353) declares that she is putting forward a defense for the claim that there is no tension between Davidson's anti-conventionalism, or his *"no-language picture"* and radical interpretation, but I fail to find such a defense, or any chance of reconciliation, in what she says next. She invokes Davidson's claim that a speaker can only intend her words to be understood in some particular way when she has good enough reasons to expect her interpreter to so understand them. She then concedes that in the case of radical interpretation the speaker must use her words consistently if she is to give her interpreter a realistic chance of getting to understand her. She then concludes that "[r]adical interpretation thus appears as a limiting case on the no-language picture: The case where all other ways of being interpretable are blocked, and the only chance at communicating lies in the regular application of one's words." (ibid.) I cannot see how we can take this last passage to amount to anything less than the full acknowledgment of the tension Glüer was trying to alleviate. Proving it a *"limiting case"* could, at most, reduce the amplitude of the clash, but not avoid it. Things look even bleaker since the tactic in this first argument for a general dispensability of conventions and regularities was to refer all communication to that fundamental moment, radical interpretation, where conventions and regularities could not be explored. Radical interpretation was supposed to act as the good exception, but what Glüer proposes is that it is the bad one. The problem is that we are still lacking a cogent case for the general dispensability of conventions in the remaining communicative settings. Maybe Glüer is not arguing with general non-necessity in mind.

speaker is conducting herself in precisely that way. We saw in Chapter 1 how the interpreter must accumulate observations to arrive at *general* statements relating the expression being interpreted with an adequate description of the *type* of circumstances that prompt assent to it.

Regularity and repetition are vital in this process. And once we grant that regularity on the part of the speaker and the correspondent knowledge on the part of the interpreter must be in place, it is relatively easy to see that the remaining conditions for there to be a convention in the Lewisian sense are also satisfied[150].

Let us simplify the community to include only the speaker and the interpreter - it is indifferent to our question if interpreter and speaker are actually establishing new conventions of their own, or if the speaker is simply introducing the interpreter into an already established and more widely shared practice. Let us use a very rudimentary example, for instance, the regularity - let us call it *R* - *of* using "There's a rabbit" to mean *there's a rabbit* - that is, the regularity of uttering "There's a rabbit" to mean *there's a rabbit*, and of taking someone else's utterance of "There's a rabbit" to mean *there's a rabbit.*

At the beginning, early on, only the speaker is conforming to *R*, while the interpreter is assuming that there is some regularity but is still trying to figure out what. He gathers that it will have something to do with rabbits, or maybe animals, or maybe game or meat. Only new observations will allow him to confirm with reasonable certainty that the right meaning is *there's a rabbit*, and, by then, conformity to *R* will already be common to interpreter and speaker. Even if the interpreter, for some reason, were not himself uttering the learned expression, he would still be conforming to *R* just by taking the speaker's utterances to mean the relevant meaning. All this, however, already gets us more than simple shared conformity. Let us see.

We can start with the question of the participants holding *common knowledge* of what is going on. The transition into competence with some linguistic expression will likely be a very diffuse event but, from the moment where there is solid enough understanding for some expression, we can be confident that interpreter and speaker will be sharing not only conformity to the relevant regularity but also common knowledge that they do. First, they

[150] Most of the time Davidson is not even that particular about whether he is talking about proper conventions in the Lewisian sense, or simply of regularities and knowledge of regularities. To show the need for regularities would already be enough to contradict a lot of Davidson's pronouncements. However, as I just said, the rest comes almost for free so why not go all the way and defend that proper conventions are needed for radical interpretation?

must know of their own case. Each participant cannot fail to be aware of the regularity *R* in his or her own conduct. Additionally, by now we also know that the interpreter knows of the regularity on the speaker's part. That was the goal of interpretation, and it is the possession of that knowledge that explains the interpreter's own conformity to *R*. Now the question is why must it be the case that the speaker reciprocates the attention? One way to defend this necessity is to invoke Davidson's criterion for communicative success - that the speaker must manage to get her words to be understood as she *intends* them to be understood. About *the Humpty Dumpty objection* (Davidson, 1986, 97-8), Davidson notes, with propriety, that intentions require a reasonable expectation of realization[151]. In our case this means that the speaker can only intend her words to be understood in a certain way when she has good reasons to expect the interpreter to understand them so. In the current context, this implies that the speaker must somehow have been able to build a justified belief that the interpreter is himself conforming to *R*. Communication will be solid only when both participants can be confident of their shared conformity to *R*. But then, since each participant is interested in what the other thinks, and all *investigation* takes place *in the open*, we cannot doubt that each participant will also know that his or her interlocutor will know things to be so, and also to know the other knows that he or she knows, and so on, until they stop *bothering*[152] to compute.

In the discussion of common knowledge, we already touched on a further article of Lewis' analysis: the condition that it should be the belief that the other conforms to the

[151] A misunderstanding of the type of view on the role of the speaker's intentions in the determination of the meaning of her words that Davidson, among others, endorses, has prompted critics to establish an affinity with Humpty Dumpty's position in his dialogue with Alice:

> *'There's glory for you!' [said Humpty Dumpty]*
> *'I don't know what you mean by "glory," Alice said.*
> *Humpty Dumpty smiled contemptuously. 'Of course you don't — till I tell you. I meant "there's a nice knock-down argument for you!"'*
> *'But "glory" doesn't mean "a nice knock-down argument,"' Alice objected.*
> *'When I use a word,' Humpty Dumpty said in rather a scornful tone, 'it means just what I choose it to mean — neither more nor less.'*

In reply, Davidson refers approvingly to Keith Donnellan's response: "Donnellan, in answer, explains that intentions are connected with expectations and that you cannot intend to accomplish something by a certain means unless you believe or expect that the means will, or at least could, lead to the desired outcome. A speaker cannot, therefore, intend to mean something by what he says unless he believes his audience will interpret his words as he intends." (Davidson, 1986, 97) This is the crucial counterweight to the insightful but easy to distort observation that speakers' intentions can determine the meaning of their words: speakers are not free to form a communicative intention regardless of their own prospects - subjectively assessed - of fulfilling them.
[152] See the first footnote in §7.1.

regularity to give everyone a good and decisive reason to conform to it. We have already noted that this is how it is with the interpreter - to figure out the regularities in the speaker's linguistic conduct so as to be able to ascribe her meanings in accordance is the proper goal of his interpretive effort. The same must also be true of the speaker; we can infer it as the best explanation for her regular conduct. It is only because she expects that her regularity will allow the interpreter, sooner or later, to adequately interpret her words that she holds on to her consistent practice. That is, it is only because she hopes and aims for the interpreter's conformity to *R* that she herself conforms to it.

The two final conditions are also fulfilled. First, once interpreter and speaker have started, there comes into place a general preference for general conformity to *R*. Second, it is also the case that there are alternatives to R that would be just as viable, as attested by the great variety of ways, in different languages, of meaning *there is a rabbit.*

Hence, I conclude, as anticipated, that radical interpretation does not provide the sort of example Davidson needs to support a claim of general dispensability of conventions for linguistic communication. Instead of a convention-free scenario, this type of situation constitutes the very enactment of a process of initiation of the interpreter into a series of conventions of language that he will end up sharing with the speaker.

7.3.3.1.1. LEPORE AND LUDWIG'S *IN PRINCIPLE* GENERAL DISPENSABILITY

Lepore and Ludwig arrive at a partially different conclusion. They affirm that radical interpretability implies *in principle* - but not *in practice* - general dispensability of conventions. They affirm that to ask "whether it is in principle possible to interpret another without appeal to prior knowledge of conventions" (Lepore & Ludwig, 2005, 278) is tantamount to asking "is there knowledge an interpreter could in principle have, leaving aside natural limitations of knowledge and perspicacity, which would enable him to correctly interpret a speaker of whom he had no prior knowledge at some given time?" (ibid.) This latter, in turn, they find tantamount to asking "whether there are facts independent of linguistic conventions that determine what a speaker means by his words" (ibid., 279). Finally they conclude that "if Davidson's basic methodological stance on matters of meaning is correct" (ibid.) - that is, if speakers are really radically interpretable in the terms set by Davidson - the answer to all these questions must be affirmative.

The whole idea behind radical interpretation is that the meanings of some speaker's utterances must be determinable by facts publicly available even to those that do not already understand her language. If we assume, as Lepore and Ludwig do, that such facts are "independent of linguistic conventions", we must conclude that if a speaker is radically interpretable he will also be interpretable without "without relying on any prior knowledge of any conventions or regularities" (ibid.).

The problem with the argument in the last paragraph, I would argue, is that we cannot assume those facts to be "independent of linguistic conventions". The meaning determining facts constitute by themselves the regularities nuclear to meaning conventions. The gathering of the remaining necessary conditions to get to full-blown Lewisian conventions is sanctioned by the kind of reasoning presented in the previous section. The condition stating that the expectation of general regularity provides each participant with a reason to conform is of special importance, what becomes evident once we consider it more closely is that those facts would not even take place were it not for those expectations.

What I am saying is that to base interpretation on the facts that make up the speaker's history of verbal behavior is not to abstain from conventions and regularities. Lepore and Ludwig arrive at the same conclusion for the question of *in practice* dispensability of conventions. As noted above, they insightfully point out that the work of the field linguist does not avoid reliance on conventions and regularities, that quite the opposite is the case, that his work consists of discovering which conventions and regularities are in place for the relevant community (see ibid., 279-80).

I do not think that the distinction between *in principle* and *in practice* dispensability can make the kind of difference they seem to think it makes. They seem to hold that regularities and conventions are only required due to practical limitations *because*, once we abstract from such contingencies, we are *simply* left with the foundational facts of speakers exhibiting particular dispositions to linguistic behavior. They seem to hold that the facts are there whether or not the interpreter is able to collect them, and that the *in principle* qualification ensures, hypothetically, that he is. If this is the reasoning, I believe it is inaccurate: to reiterate the point made in the last paragraph, the argument fails to recognize that these facts in themselves are not "independent of linguistic conventions".

To be sure, however, the type of special powers suggested by the qualification allow for the construction of farfetched scenarios where there could, perhaps, be communication

without the perfect fulfillment of Lewis' conditions for convention. In any case, too much would still be in place for these fictions to constitute any acceptable vindication of the strong non-necessity claim. Here is one such story: imagine that speaker and interpreter are for the first time face to face. The speaker utters "There is a rabbit" and the interpreter, (somehow) fully and instantaneously informed of the speaker's history of uses of the expression, correctly figures out what she means. Here, we still have regularities on the speaker's part - in her history of uses of those and other expressions. But we may perhaps release the interpreter from any regularity. The regularity on the interpreter's part would ordinarily be needed to support the speaker's expectation of being interpreted to mean what she intends to mean. But in this new case it seems possible that the speaker will be able to ground her reasonable belief in some other way - perhaps she would be aware of the interpreter's special powers. One could still continue to challenge the point, in particular, with some rejoinder based on the observation that, even if each particular interpreter may avoid regularity by interpreting the speaker's utterance of "There is a rabbit" only once, we still need some regularity in the interpretation *across interpreters* to justify the speaker's consistency. But I will stop here. I conclude with two thoughts: on the one hand, I find it very dubious that there really is anything to be learned about verbal communication and what is essential to it from such exotic scenarios as this; on the other, after a first timid incursion it appears that not even in such *wildernesses* can we picture verbal communication taking place without regularities and much of the stuff of Lewisian conventions.

7.3.3.2. THE ARGUMENT FROM THE ESSENCE OF COMMUNICATION

The second argument consists of two steps. First, it stresses how being understood as one intends to be understood is essential to linguistic understanding - a sufficient and necessary condition for successful communication, its ultimate and most direct criterion. Then, once the centrality of the communicative intention and its recognition by the audience have been emphasized, all else is put aside as secondary and merely contingent.

Where the argument seems more clearly present - even if, perhaps, not fully articulated - is in Davidson's 1994 *The Social Aspect of Language*. It turns up almost explicitly in the following passage:

> … so far as the point of language is concerned, our only obligation, if that is the word, is to speak in such a way as to accomplish our purpose by being understood as we expect and intend. It is an accident, though a likely one, if this requires that we speak as others in our community do. (Davidson, 1994b, 118)

Besides, several times Davidson declares the essentiality of communicative intentions and their recognition for communication:

> … what matters, the point of language or speech or whatever you want to call it, is communication, getting across to someone else what you have in mind by means of words that they interpret (understand) as you want them to. (...)
>
> The intention to be taken to mean what one wants to be taken to mean is, it seems to me, so clearly the only aim that is common to all verbal behavior that it is hard for me to see how anyone can deny it. (ibid., 120)

And several times this appears contrasted with the incidental character of the presence of regularities and conventions in the affair:

> I have stressed both the obvious utility of the large degrees of overlap in verbal performance we find in groups that live and talk together, and the inevitability that conformity will be learned and encouraged. The theoretical possibility of communication without shared practices remains philosophically important because it shows that such sharing cannot be an essential constituent in meaning and communication. (ibid., 119)

I decidedly agree with everything in the first step, but what I question is the second one, the idea, at least suggested, that the essentiality of some aspect implies the mere contingency of all others. More precisely, what I question is the idea that if language is essentially an instrument of communication, it cannot be essentially regular or conventional, since - this seems to be the reasoning behind it - if success in communication is first and foremost a matter of having communicative intentions adequately recognized, any way to bring this about will be equally appropriate, with regularities and conventions being simply *one means among others*.

The inference in the second step, I defend, is incorrect. Conformity to regularities and conventions, I agree, is not the point when we use language. It does not constitute a direct or definitive criterion of success for verbal intercourse, it plays different types of roles and is accomplished to different degrees in different instances of sound communication, and its connection to the ultimate goal of verbal understanding is more intricate and loose than one

might be tempted at first to suppose. Even so, none of these prevent regularities and conventions from playing some form of indispensable role in every communicative episode, if not by supporting standard and codified interactions among speakers to occur in the most fluent and predictable way, then, at least, by providing the background that makes possible semantic innovation and ingenuity. Indeed, having now reflected on the available arguments to the contrary, this is still what appears to be the case. I think the onus is still on the *radical anti-conventionalist*, if there really is one, to prove this to be wrong, and regularities and conventions to be generally dispensable.

There is no opposition or exclusivity between linguistic communication being essentially an exercise in the production and recognition of communicative intentions and its also being, to some extent, necessarily regular and conventional. The realization that the right sort of transaction in communicative intentions is what we need, and all we need, for communication does not further the case against the necessity of conventions and regularities. It does not allow for such a consequence since, with no telepathy at hand, all available examples of successful transactions in communicative intentions appear still to rely, directly or indirectly, on those conventions and regularities.

7.4. CONCLUSION

The arguments considered in the previous sections lead me to conclude in favor of the general insufficiency and the merely particular unnecessariness of meaning conventions for linguistic communication.

Accordingly, the interpreter cannot adopt reliance on meaning conventions and knowledge of the standard use of expressions in a community as his sole method of linguistic interpretation. I argued that the ever-present possibility of words not being used in accordance with their standard or expected meaning requires the interpreter to engage in an ongoing effort of general interpretation that will have the last word in his interpretive decision.

On the other side of Davidson's anti-conventionalism, the results are more modest. I even rest uncertain about whether to ascribe to Davidson the endorsement of a strong and general claim of non-necessity. Although various of his pronouncements appear to affirm exactly that - above I identify some passages - the fact is that he does not unmistakably

engage in any argument clearly aimed at supporting the strong thesis. There seems always to be some oscillation between the strong and general and the weak and local claim of dispensability. But be that as it may, the fact is that we found no way for verbal communication to go on in the complete absence of regularities and conventions. Occasionally, verbal communication will be successful without being directly sanctioned by prevailing meaning conventions, but from this one should not extrapolate too much. I considered two lines of argument, *the argument from radical interpretability* and *the argument from the essence of communication* and defended that they are ineffective in securing general dispensability. One is even reminded of Davidson's insight on a *non sequitur* he revealed important to bear in mind when considering the possibility of massive error. The topics are quite separate and the claims at stake are absolutely independent, but there is an interesting analogy of form. Davidson stresses that from the fact that no belief is safely guarded from the possibility of error, it does not follow that all beliefs could massively be wrong[153]. In an analogous way, in our present case, I defend the idea that we must refrain from taking the likely fact that "there is no word or construction that cannot be converted to a new use by an ingenious or ignorant speaker" (Davidson, 1986, 100) to imply that all words or constructions could be given a new use every time. In other words, from the fact that every meaning convention could be subverted, it does not follow that all conventions could be massively relinquished.

This failure in securing strong non-necessity is no considerable drawback in the defense of my main thesis. The results summed up above are enough to support the inadequacy of *the stiff conception of language*, and to confirm the need, advocated by Davidson, and the central object of this dissertation, of investigating into the speakers' minds to arrive at each interpretive judgment in every instance of verbal understanding. This investigation calls for more than the almost automatic execution of a stable method of utterance decoding. This investigation involves the interpreter in a complex and inventive exercise, one that mobilizes an undetermined collection of information and capacities, and aims at the somewhat vague end of making overall sense of speakers' conducts. In sum, this investigation requires rationalization. Certainly, the intensity and depth of the effort will vary from situation to situation, as well as the interpreter's awareness of it, but in one form or

[153] See, in particular, Davidson, 1983.

another, there will always be rationalization - at least, this chapter's discussion adds further support to our thinking so.

8. Radical Conventionalism: Lepore and Stone on Meaning Conventions

In this chapter, I consider and resist a radical form of conventionalism characterized by the endorsement of the claim that all that properly belongs to meaning and verbal communication is fully determined by conventions and standards. Ernest Lepore and Matthew Stone (L&S) substantiate this position in their recent 2015 book, *Imagination and Convention*[154].

I start in §8.1 with a presentation of L&S's proposal, *direct intentionalism*, and the characterization of some central theoretical elements employed in it. These later include an original species of intention, and a number of institutions and mechanisms with which L&S hope to constitute the sort of action and the sort of social environment that would - without further considerations regarding the speakers' own desires, expectations and communicative intentions - commit speakers to the community meaning of the expressions they use, thus paving the way for a form of interpretation that is potentially automatic and dispenses mindreading and rationalization.

In §8.2, after the stage is set, I place my focus on L&S's direct intentionalism's response to the especially troublesome cases of malapropisms and other *faulty* utterances. I oppose their approach to such cases on two scores.

First, I find them too prompt in committing the speaker with a meaning she does not desire, believe or intend to mean. In §8.2.1, I consider the two lines of argument explored by L&S in favor of an inflexible strategy of linguistic interpretation in accordance with the community standards: L&S invoke both Lewisian conventions and externalist institutions in explaining the alleged commitment to the community norm; in §8.1.2.3. I explain why I do not count L&S's appeals to the conversational record as a third independent argument.

[154] Ernest Lepore has kindly informed me of some very recent developments in his and Matthew Stone's position on some of the topics discussed in this chapter. I learned about this just a few days before submission of this dissertation was due, and thus too late for any rectification to be possible. This note has the purpose of warning the reader of the discrepancy.

Taking each in turn, I argue - in §8.2.1.1. and §8.2.1.2, respectively - that neither conventions nor externalist institutions can allow a speaker to mean the standard meaning of any expressions in the absence of the relevant supporting mental states. In general, I conclude, there is no genuine linguistic understanding among interlocutors without a corresponding minimal meeting of minds. A successful meaning ascription must capture what the speaker expects to mean. Anything else would, to some extent, obstruct understanding and damage conversation.

Second, I find them too cautious in not allowing the meaning actually intended by the speaker for her words and adequately recognized and handled by the interpreter to be acknowledged a semantic status. In §8.2.2, I defend that the notion of intended, potentially non-conventional and non-standard, linguistic meaning that would allow us to take those benign cases of non-standard uses of words that still do not cancel mutual understanding as genuine cases of linguistic meaning and communication is needed also to play many other roles essential for linguistic communication and the social uses of the language as we know them. I argue that we already need that notion to explain what motivates speakers in their social practice of the language, to account for the emergence and consolidation of linguistic conventions and standards, to understand the possibility of ephemeral deviations from such uses, and to account for word learning, both at the early stages of language acquisition and later on, since speakers never cease to be offered opportunities to improve on their linguistic knowledge. I defend the viability of a conception of language and interpretation based on such a notion of meaning against a number of objections raised by L&S.

I use the specific cases of faulty utterances to make vivid a perfectly general disagreement between L&S's radical conventionalist approach and the conception of meaning that I endorse and have been developing in this dissertation - the conception that underlies the claim that linguistic interpretation always involves rationalization. In the end, it is the same two theses that emerged in the previous chapter - the general insufficiency and the occasional non-necessity of conventions for linguistic communication - that are at stake here. In this chapter, I continue the defense started there, but this time against the specific difficulties and radically alternative proposals put forward, very recently, by L&S.

8.1. LEPORE AND STONE'S *IMAGINATION AND CONVENTION*

As neatly reflected in their title, *Imagination and Convention*, L&S argue that the prime distinction to be drawn in the study of language and of verbal communication is the one that separates conventional practices on one side, and imaginative exercises on the other, with the ponderous consequence of abolishing an intermediate level between them and, with it, in particular, the paradigmatic pragmatic category of conversational implicature that has application only there.

L&S extend the scope of the conventional far beyond the traditional confines of semantics - see, especially, Part II of their book. For an important collection of cases - cases involving presupposition, anaphora, indirect speech, and information structure - where the received view detects the work of conversational implicatures in the generation of successful communication, L&S argue that what we actually find, if we look carefully, are simply more linguistic conventions; where the orthodoxy depicts speakers engaged in complex and general purpose pragmatic reasoning, they see them, instead, simply making regular use of their acquired knowledge of their particular language.

Pragmatics is reduced to a bare minimum - it "merely disambiguates; pragmatic reasoning never contributes content to utterances" (Lepore & Stone, 2015, 83). There is nothing more, in their words, "[t]here are no special meanings, over and above the meanings of our utterances, that interlocutors infer by calculation from a Cooperative Principle, maxims of conversation, or other general principles for pragmatic enrichment and reinterpretation." (ibid., 199)

More generally still, they affirm that "it is a mistake (...) to think that "mind reading" inevitably intervenes in meaning-making"(ibid., 6) or that "listeners normally recover the meanings of utterances by recognizing the conventions involved, not by reasoning about the speaker in any deeper sense" (ibid., 199). I find these qualified assertions - by "qualified" I am referring to the use of "inevitably" and the "normally" - actually misleadingly modest in face of the proposals they then go on to elaborate, where indeed they leave no room at all for *deep* inference, reasoning and mindreading in the process of semantic interpretation, and do not allow speakers to generate genuinely linguistic meaning outside the standards and conventions in force in the community.

Outside linguistic conventions there is only the imaginative, creative and free forms of engagement with language, which do not commit speakers with fixed meanings and bring

about essentially unpredictable effects in the audience - see Part III of the book. On this part of the divide, they explicitly list suggestions and hints, metaphors, irony, sarcasm and humor.

In sum, here is the radical conventionalism of their proposal: speakers can only generate meaning in accordance with conventions and prevailing standards; unregulated steps or effects involving language, whether or not they result in genuine understanding among agents, are deemed foreign to meaning proper.

Taking issue with this view's viability means returning to the hypotheses, already discussed in the previous chapter, of conventions and standards being sufficient and necessary for linguistic interpretation. The arguments deployed here continue the line initiated there, elaborating on aspects already noted, but expand also in new directions.

In reaction to L&S's book, a fair number of commentators reported some form of reluctance in accepting the claim that all communicated meaning in language - all coordination among interlocutors on the content of some utterance, all recorded contributions to an ongoing conversation - is conventional[155]. They mostly focus their critique and arguments on what pertains to the traditional domain of pragmatics. Almost all of them concede that L&S are right in claiming that there is more conventionality to language than is usually admitted, but still perceive their proposed revision as too drastic. In one form or another, they all defend the necessity to acknowledge various ways in which the meaning exchanged on particular occasions of linguistic intercourse exceeds the conventional, or otherwise fixed in advance, meaning of the expressions used.

They raise their doubts mostly concerning the need to account for what comes after a first fixed layer of meaning, where issues such as implicatures, explicatures and semantic underdetermination have been feeding intense inquiry and discussion for decades. In contrast, my own questions concern that first layer[156], orthodoxically associated with what is usually referred to as linguistic meaning, or expression meaning, as opposed to speaker meaning. My main contention here is not that the interpreter needs to go beyond convention in order to capture the subsequent layers of content that are pragmatically added to that initial base. My main contention here is that, simply to earn the ascription of that first level of meaning, the

[155] See Bezuidenhout, 2016, Horn, 2016, Szabó, 2016, Carston, 2016, and Collins, 2016.
[156] As much as possible, I am adopting a neutral position concerning what should be included in this first level, and whether, for instance, it is enough to fix truth-conditions, falls short of that, or goes beyond it, into the determination of other aspects of a speaker's contribution to a conversation.

interpreter cannot merely rest with his knowledge of conventions and norms in the community of reference. He is also required to engage in mindreading and rationalization.

8.1.1. DIRECT INTENTIONALISM VS. PROSPECTIVE INTENTIONALISM

L&S name their own view "*direct intentionalism",* and mean to contrast it with *prospective intentionalism*, the label under which they include all broadly Gricean conceptions of verbal communication.

Prospective Intentionalism is a broad category intended to lodge a large number of positions and proposals. What all those positions and proposals have in common is their reserving a robust and convention-independent role for intentions in the determination of meaning. The type and extent of meaning that is intentionally determined may vary from case to case. We may be talking only of pragmatic inferences built upon a stable conventional ground, or we may be talking about the whole meaning of an utterance in extreme cases where linguistic understanding is attained in an improvised way not based on previous knowledge of the language. The common element in all relevant cases is that some content is exchanged among interlocutors - that is, a content is specified in the communicative intention behind the speaker's utterance, and adequately recognized and handled by the interpreter - in ways that transcend reliance on meaning conventions or any other type of pre-established forms.

Grice, the paradigmatic case of Prospective Intentionalism, makes room for all these degrees of autonomy of intention in meaning making. He naturally acknowledges the fixed and conventional character of public languages but, in taking utterer's meaning as the primitive notion in his semantic edifice, he accepts as genuine cases of communication all situations in which interlocutors manage to fulfill the set of intention-based conditions he refined over the years in his analysis of utterer's meaning. He seems thus ready to accommodate even cases of improvised understanding over non-standard uses of words such as the malapropisms discussed in the previous chapter.

The position I am defending in this dissertation is flagrantly within the range of Prospective Intentionalism. First, as explained, in defending a central place for rationalization in interpretation, I am defending that the use of language must be understood in the bigger context of human interests and activities. I endorse the view of linguistic communication as a

piece or aspect of a general enterprise of understanding rational agents, thinkers and speakers, and I believe that interpretation cannot take place without taking into account the continuity between agents' linguistic behavior and the rest of their human and intelligible life. As I am convinced that it is agents' general understanding of each other that simultaneously allows, motivates and determines linguistic practices, I cannot accept conformity to meaning conventions or to the standard practices in the community to be taken as ends in themselves or as definitive criteria of success in the use of language. They are mere instruments and stages in a more fundamental and wider process or enterprise.

Second, I also follow Grice's insight about the essential role of communicative intentions and their recognition in meaning transactions. Furthermore, and taking into account what is summed up in the previous paragraph, I believe that the relevant intentions cannot be so local or shallow as to fail to be rationally integrated in the speaker's mind. This acknowledgement of the role of intentions in meaning making is patent throughout what precedes this chapter, and highlighted in particular in Chapters 7 and 4.

Direct Intentionalism is the novelty here and will require lengthier characterization. In going with "intentionalism" for their brand's name, L&S signal that, just like everyone in the Gricean lineage, they reserve a central place in their account for *intentions*, namely, *speaker's intentions*. However, they swiftly detach themselves from that tradition. They "recommend a more modest, but superficially similar, view: That the speaker's intentions determine the meaning of an utterance by linking it up with the relevant conventions." (Lepore and Stone, 2015, 200) What distinguishes their proposal is precisely the conviction that speaker's intentions cannot determine meanings independently of previously established linguistic conventions.

L&S's choice of "direct" is an explicit acknowledgement of the inspiration they take from "Kripke's theory of direct reference" (ibid.) and begins to illuminate the particular understanding of the social structures of language they have in mind. L&S trace their conception of *convention* to Lewis' analysis of the notion - presented in the previous chapter - but later in the book they significantly qualify their endorsement of his account. In fact, their emphasis on the conventional appears at times more congenial with certain positions characteristic of what is usually referred to as *Semantic Externalism* than with Lewis' reflections on the expedients by means of which two rational agents may concoct a solution to a coordination problem. Since we are here, I must not fail to mention yet a third type of

social device that L&S invoke in their account, the *conversational record:* "a definitive, precise, and circumscribed inventory of the contributions that interlocutors have made to the conversation." (ibid., 234). Later in this chapter we will have the chance to more carefully inspect each of these items.

8.1.1.1. THREE TYPES OF INTENTIONS

A proper elucidation of each brand of intentionalism and of the contrasts between them - as L&S conceive them - involves an elaborate taxonomy of intentions. L&S distinguish three general types of intention. They make use of the distinction between *future-oriented intentions* and *intentions in action* - familiar at least since Searle (1983) - and then they add a third type of intention to the lot, *basic intentions*, to which they concede an essential role in their view.

We can order the three types with respect to their temporal and logical proximity to action, with future-oriented intentions being the farthest from action, basic intentions the closest, and intentions in action taking their place at an intermediate level.

Future-oriented intentions correspond to broad plans of action, still to detail as the relevant occasions arise. The agent's "commitment is something of an abstraction; the course of action [she] envisage[s] is incomplete and only partially specified" (Lepore & Stone, 2015, 207). Future-oriented intentions will direct her conduct but they are unlikely to determine, from the start, a very precise course of action. Pursuing or acting on such an intention "means getting these details worked out" (ibid.). As an example they offer vacation planning, "going to Provence for August" (ibid., 207), and explain how "this intention sets up a whole set of sub-problems" (ibid.): "You have to buy plane tickets and arrange a place to stay; you have to arrange for someone to feed your cats or water the plants, and so forth" (ibid.).

Intentions in action are the intentions presiding over the agent's action, but they are still on the *conceptual side* of the deed, so to speak. Intentions in action are specified by "what someone was trying to do or had in mind when she undertook a particular action" (ibid., 210) Here are included all effects that the agent, when acting, anticipated and committed to, and any description that is recognizable by the agent as fitting her action should do. L&S use the hanging of a picture as an example, to which a number of intentions in action are likely associated. As they explain: "any one of the agent's commitments

underwrites its own way of describing what [she is] doing as [she] swing[s] the hammer: hitting in the nail, affixing the hook to the wall, anchoring the support for the picture. These are all intentions in action" (ibid.).

L&S find room for yet a third type of intention - basic intentions. Just like intentions in action, basic intentions are contemporary with the action. Their distinctive feature is their more concrete and immediate character, which is supposed to allow them a direct control of action[157]. According to L&S's proposal, it is intentions of this type that are actually able to terminate a deliberative process and to mediate between thought and behavior,

> We imagine the agent arranging behavior through a tightly coupled cycle of deliberation and action. Each step of deliberation infers a judgment that the agent should do a particular action. This judgment is an inference drawn in an appropriate way from the agent's beliefs, desires, and intentions. At the same time, this judgment forms the basis for the agent's actually carrying that action out—with each judgment at each step of action, the agent decides to do something, and then straightaway does it.
>
> These more fundamental judgments are what we are calling basic intentions. They mediate between our represented knowledge and preferences and our capacities for action. (ibid., 208)

Basic intentions are meant to engage only "the agent's fundamental capabilities" (ibid., 208). L&S briefly invoke Gibson's notion of *affordance*[158], which they characterize as "effects that agents can bring about just in virtue of the kind of being that they are and the kind of engagement they have with the world" (ibid., 208), to determine the scope of actions targeted by these intentions. They offer the intention *to grasp that thing* as the prototypical basic intention. They also mention the hammering of a nail (ibid.,210).

L&S stress the presence of an element of indexicality in basic intentions. They explain that:

> In a basic intention, agents have to have direct access to any parameters of the action. In particular, the objects of actions in basic intentions must be represented indexically; like expressions in natural language such as *this* and *that*, they refer not through a name or description but through the interaction that the agent has with them… (ibid., 209)

This is intended to allow agents some form of unmediated, non-descriptive or non-conceptualized relation to the contents and objects of their intentions and actions. As I

[157] L&S's emphasis on this point seems to imply that they take intentions in action to fall short of such capacity.
[158] Gibson, 1979

understand it, this will be crucial in allowing Direct Intentionalism to make room for cases where a speaker intentionally means some unintended meaning - I will return to this topic later.

8.1.1.2. MEANING DETERMINING INTENTIONS

The proposed taxonomy allows a clearer differentiation between the two views. Here is how they affirm their distinctive proposal:

> Griceans, Relevance Theorists, and many others take these considerations to establish a fundamental link between meaning in language and what we have called communicative intentions in action. But … direct intentionalism offers another explanation. The idea is that the intention determines what you said: What words you used, what syntactic analysis you had in mind, and so forth. You had to represent the utterance to say it, that's all. ... We think there are lots of good reasons, conceptual, empirical, and theoretical, for attributing meaning in utterances in a much more direct way than Griceans do. (ibid., 225)

While intentions and intention recognition are similarly indispensable in both brands, each intentionalism attributes the determination of meaning to a different type of intention. The Direct Intentionalist allows it only to the most *superficial*, closest to action, type of intention: basic intentions. As L&S put it: "[o]nce the speaker commits to the grammar of her utterance, she has established a basic intention that links that utterance to its meaning. After that, it's not up for grabs what the utterance means." (ibid., 219) The Prospective Intentionalist holds that meaning through language involves also the other two *deeper* levels of intention - that meanings are ultimately determined by speakers' intentions in communicative action - and of the larger and richer portion of the speakers' minds needed to accommodate such intentions. All differences diagnosed by L&S between the two forms of intentionalisms either follow from or elaborate on this fundamental divergence concerning which type of intention should be given the central role in linguistic communication.

The most immediate consequences of this divergence are manifest at two levels. First, Prospective Intentionalism will be committed to a stronger form of integration of the meaning determining intentions within the speaker's mind. In contrast, the Direct Intentionalist allows for some amount of disruption. Meaning is taken to be determined quite locally by basic intentions that - presumably, because of the indexicality of their contents - can be at odds

with much of what is going on in the speaker's mind. We will look at this more closely in cases of malapropism.

Second, while the Prospective Intentionalist leaves the speaker relatively free to choose her way to make her meaning intentions known to her audience, the Direct Intentionalist constrains the speaker to do so by means of some pre-established system, a public language in which she counts as competent. This difference in the mechanisms of meaning making is crucial to L&S's plan of providing a sharp distinction between genuine transactions in linguistic meaning and the broader issue of general understanding among humans.

The proposal is that an agent's competence in some language constitutes a particular type of affordance in her repertoire. To be a speaker of a language is to hold the "fundamental capability" of "contributing [a] grammatically specified meaning to an ongoing conversation" (ibid., 208) by "performing an utterance of a specified linguistic structure" (ibid.). To exercise that capability the speaker needs only a basic intention. To *say something in the language* becomes on a par with *hammering the nail*, *pushing the button* or *grasping something* in terms of psychological immediacy. That is, in this view, the deeper layers of beliefs, desires, plans and deliberation are not decisive in the specification of the speaker's linguistic performance. *With the right conditions in place*, the action is fully determined by the speaker's basic intention to utter an expression with some specific linguistic form just like an act of *nail hammering, thing grasping* and *button pushing* can be accomplished with a certain basic autonomy from what else may be going on in the agent's life and thoughts.

Now, what is it for *the right conditions to be in place*? In L&S's view, speakers manage to use their language in communicative interchanges in virtue of what they refer to by means of various descriptions such as "the environment that constitutes meaning" (ibid., 219), "the operative context" (ibid.), or even "the general background conditions for using language meaningfully" (ibid., 222) - henceforth just "environment". What is this *environment*, and how is it constituted and brought about? What is supposed to allow speakers the type of direct access to the standard meanings of expressions in their languages described in the previous paragraph? As already noted above, L&S identify three types of social structures doing some of that work: conventions, semantic externalism institutions, and the conversational record.

8.1.1.2.1. CONVENTIONS

The pervasive presence of linguistic conventions in our use of language is the central theme of their book. L&S offer a careful elucidation of the notion with which they are working. They claim to be employing Lewis' notion of convention, or at least to take it as a starting point. Lewis' analysis was presented in the previous chapter and there is little point in recapitulating those fundamentals here.

Later in the text, L&S qualify their endorsement of Lewis' proposal and identify some of its shortcomings. They point, in particular, to the proposal's inability to cope with the dynamic character of language (in §14.3.1) and to what they see as its inadequate fit with the existence of a Chomskyan Universal Grammar, or other innate, non-learned aspects of our linguistic competence (in §14.3.2).

These late divergences still seem patently insufficient to explain or accommodate L&S's occasional invocation of the notion at times where, according to the Lewisian conception, it seems to have no application - in particular, cases where interlocutors do not actually share sound knowledge regarding the regular use of some expression in the community. I have one particular type of example in mind - malapropisms. More to that point, L&S also express their discontent with Lewis' affinities with Grice and Prospective Intentionalism (ibid., 233-4, 241). The concern is that Lewis' notion and analysis leads to an account of conventional communication that depends too heavily on the verification of a certain type of collaborative interaction, as well as of a certain degree of familiarity, among interlocutors, with each other's mental states that L&S find ultimately dispensable. However, it is hard to imagine what could survive of Lewis' proposals and insights if one were to extirpate this Griceanism from it. I do not think L&S actually manage to solve the impasse. An appeal to conventions cannot help in overcoming Griceanism if Griceanism is built into our notion of convention. This tension will come to the fore when, in a moment, we get to the discussion of L&S's position on the meaning of malapropistic utterances.

A second crucial notion that L&S pick from Lewis' work on convention is that of *coordination* - and they even follow it outside conventions proper, as we will see below with regard to the conversational record.

As presented above, Lewis defines conventions as rational solutions to coordination problems. These are situations where two or more participants share a common goal but can

only attain it if they manage to match their strategies. Success is dependent upon the agents' ability to concert their efforts and this, in turn, is dependent upon their ability to form adequate expectations regarding each other's conducts. L&S identify *coordination*'s great potential as a notion upon which to base the distinction they are trying to articulate: genuine verbal communication to one side, with effective coordination in the sharing of contents, clear commitments on the part of participants, and determined contributions to the conversation and interpretations; creative, less constrained and less concerted uses of language to the other.

The crucial detail is that they seem unwilling to accept the possibility of coordinated strategies in linguistic interaction without some form of prearrangement, however realized and described - for instance, a "learned expectation" (ibid., 240), or a "social competence shared among interlocutors" (ibid., 216). Be it in the form of an already established use of the language and, hence, directly instrumental in the securing of linguistic understanding, or, ultimately, in the form of an established procedure for the invention and establishment of new uses[159], L&S always insist on some form of pre-established agreement. Here is a passage where, if I understand them correctly, they state precisely this:

> Our theory of semantics has to be compatible with universals of meaning, with improvised meanings, and with our practices for disambiguating, refining, and negotiating meanings. These aspects of meaning do not fit Lewis's understanding of convention. These aspects of meaning are still manifestations of interlocutors' social competence, we will suggest—but *it involves knowledge and mechanisms of a specific kind.* In particular, interlocutors are normally coordinating on a process of inquiry, through which they commit to make their meanings public. *Semantics*, we will suggest, *describes the social competence that specifically supports this coordinated inquiry*. (ibid., 245; emphasis is mine)

I concur on the idea that the distinction between coordination and lack of it is a very basic and important one[160] in the study of language and its uses, but I do not follow L&S in sustaining that the only way to secure coordination in verbal communication is by means of

[159]In their own words, "if much of meaning is in fact improvised, then we will always need to describe institutions that let us make meaning, not just institutions that let us use established meanings" (ibid., 243).

[160] Whether or not it should be used to delimit the scope of semantics, as L&S claim it should, is another question on which I take no position.

prearranged solutions associated with specific "social competence[s]" that dispense agents from unscripted and eclectic reasoning and inquiry, from mindreading and rationalization[161].

8.1.1.2.2. EXTERNALISM'S INSTITUTIONS

Second in the list of social structures with a plausible role in the constitution of the *environment,* we have the machinery emblematic of the "externalist view of meaning, grounded through a division of linguistic labor" (255) that L&S endorse: the "network[s] of causal and informational relationships that ultimately connect [speakers] with the word's meaning" (209). The distinction between the first and the second type of structure is perhaps subtle and, at times, understated and underappreciated. In passages such as the following, L&S still formulate their doctrine in terms of conventions, but what really seems to be operative there are not the Lewisian notions and insights, but the externalists' ones.

> The general assumption is that, because we subscribe consistently to linguistic conventions, they apply to what we say, as long as some very general background conditions are met. A word, in virtue of the conventions, gives us a connection to its meaning. We can access the meaning by saying the word – it's part of our basic repertoire for interacting with each other. (ibid., 212)

In keeping with the terminological option announced in the previous chapter of this dissertation (§7.2., fn.5), I will mark the important difference between what is at stake in each stance by talking of *standard meanings* and *standard uses of expressions* when in the context of externalism, and reserving *convention* and *conventional* for when dealing with Lewisian topics and concerns.

What is the externalist machinery*?* L&S refer to Kripke (1972) and Putnam (1975) who famously uncovered the role of structures, such as historical links and mechanisms of deference at work in the linguistic practice, that free speakers from the need to earn their competence with a term or name, the grasping of its meaning or reference, only via some form of rigorous knowledge of the subject or object. That is achieved, for instance, by allowing speakers to resort to experts in determining what belongs to the extension of a certain term; or by counting them as apt referrers simply in virtue of their having borrowed the name in question from a fellow apt referrer who is previous in a chain ultimately linking

[161] More on meaning improvisation in §8.2.2.3.

back to the referred object itself.

These structures constrain and determine linguistic practice and promote the preservation of standard meanings even in the absence of perfect knowledge on the part of the speaker. Even when a speaker's knowledge of Aristotle is far from enough to allow her, for instance, to distinguish Aristotle from other persons in his vicinity, the simple fact that the speaker acquired the name in an adequate fashion may be sufficient to endow its utterance with the name's standard reference. Even when a speaker cannot tell the difference between an elm and a beech, division of linguistic labor and implicit deference to experts allows her to talk distinctively about each species of tree.

I readily acknowledge the importance and likely truth of externalism's central claims, but I am convinced that it cannot be made to do all the work L&S seem to ask of it. I will argue that Direct Intentionalism's extreme form of reliance on ideas of linguistic community and of a public and shared language, and the liberality with which it commits speakers to the standard meanings of expressions, are not sanctioned by the most familiar elaborations of the externalist doctrine.

8.1.1.2.3. THE CONVERSATIONAL RECORD

Lastly, there is the *conversational record,* strongly inspired in Lewis' *conversational scoreboard* (Lewis, 1979). It is a versatile device properly introduced by L&S only in the last chapter of the book, and is intended to allow them to conclude the theoretical consolidation of the diverse proposals and positions that came before.

The conversational record is characterized "as a definitive, precise, and circumscribed inventory of the contributions that interlocutors have made to the conversation" (Lepore & Stone, 2015, 234). It is a "theoretical construct" (ibid., 234), the complex result of a complex practice constituted by an assortment of simpler practices in interpretation and conversation, unified by the fact that all of them involve some social competence and, when successfully exercised, generate coordination between interlocutors.

The appeal to a conversational record is motivated by the need to suppress detected insufficiencies in the explanatory power of the notion of convention - especially its lack of resources in accounting for fluid and evanescent aspects of linguistic interchanges such as those, for instance, related to discourse or conversation dynamics and to moves and practices

of *disambiguation*, *refinement, negotiation* or even *improvisation of* meaning (ibid., 245)[162]. Fluidity and evanescence do not prevent interlocutors from committed, concerted and reliably informed conversations. That this is so indicates that the relevant practices and moves are somehow tracked, registered and shared among interlocutors. The conversational record is the theoretical object where all this tracking, registering and sharing is going on. It concentrates all the "diverse and heterogeneous" (ibid., 249) linguistic ingredients upon which interlocutors coordinate the fluid and evanescent along with the more stable core of conventional or standard linguistic knowledge. Only the imaginative effects are left outside. According to L&S, "the conversational record will be as rich and varied as the grammar itself" (ibid., 249); put another way, their "view is that semantics describes interlocutors' social competence in coordinating on the conversational record" (ibid., 256) - this is how, by rounding up the diverse elements collected on the first side of their divide, the notion of conversational record is meant to allow for a better consummation of the distinction that gives L&S's book its title.

Now, the thing with the conversational record is that it does not explain how anything gets to belong to it. Unlike appeals to conventions, or to the causal and informational networks of externalism, that do, in some measure, help explain how semantic facts or rules may be brought to public knowledge and fruition, the conversational record can at most be invoked to *state* the publicness of such facts and rules, not to *explain* how it has been reached. I think that this is part of what L&S are getting at when they assert "[t]he chief worry about the conversational record is that it is a theoretical construct. It offers us the freedom to recapitulate distinctions, rather than to explain them" (ibid., 234). One can still acknowledge the importance of the notion in allowing L&S to find some natural or intuitive unity in a number of otherwise scattered elements of the practice of language, but that seems to exhaust its usefulness. That is, the invocation of the conversational record appears simply meant to allow the reiteration and realignment of previous conclusions and positions, not their defense or elucidation. In particular, it seems incapable of helping us determine what

[162] L&S sum it up in this passage: "Change is crucial to all these aspects of the conversational record. The questions we are working to answer will change as we resolve some and give up on others. The list of the propositions we have agreed to will grow and perhaps shrink. Our interest will shift from old entities to new ones. We can also shift the perspective and point of view we use to present the discourse and relate to one another. And our meanings themselves can be refined. We can name things in new ways, or come to classify the world in terms of new properties. And we can draw different boundaries, for example, by sharpening our terms and strengthening our standards." (ibid., 249)

linguistic practices belong on each side of the divide because it tells us nothing about the ways in which speakers manage to coordinate or what they manage to coordinate. Thus, any appeal to this notion seems sterile to our central question here. It provides no contribution to an analysis of the *environment* that is revealing of how it can happen that speakers have direct access to the standard meanings of expressions in the way Direct Intentionalism claims they have.

It should also be added that the point made above with respect to convention and coordination is again pertinent here. L&S seem convinced that interlocutors could only get something to be *coordinated into the record* by means of prearranged forms and solutions. This is fundamental for their positions and conclusions and is made explicit in passages such as the one quoted above (end of §8.1.2.1). It is also, I believe, the main source of our disagreement.

In conclusion, to our question concerning the nature and constitution of the *environment* in L&S's proposal, two solid answers have emerged: conventions and externalism's structures and institutions. In the end, we must conclude that the conversational record does not really have a place here. It does not play the same role as conventions and externalism's institutions and structures in L&S's plan. In fact, they seem to belong in different categories, with the first two being tributaries or constituents of the latter.

To these two we must still add all the innate *knowledge* of the language, whatever exactly that may be. L&S do not adopt an elaborated position on this particular question. I share their acknowledgment of the more than likely presence of "arbitrary psychological mechanisms" (ibid., 243) positively constraining the pursuit of communicative coordination among speakers and, just like them, I will not elaborate on the topic. The focus will continue to be placed exclusively on the social (or simply interpersonal) element. This is perfectly recommendable since our interest lies not with the universals of meaning but with the specific properties that distinguish each language from all others and involve some social realization on top of speakers' biological endowments. What we want to know is how the necessary additional constraining of the range of possibilities left open by something like a Chomskyan Universal Grammar takes place. What we want to know is how we get from there to the situation where a particular speaker can use some particular linguistic form to mean some particular meaning.

We now have all the elements of L&S's alternative to Prospective Intentionalism and the conception of language and language use that I endorse in this dissertation. L&S propose a much more direct relation of speakers and interpreters to the meanings of the utterances. They defend that speakers of the same public language share a specific environment that sustains the possibility of such direct access. I have counted two effective mechanisms they put forward as constitutive of that environment. What I will do next is to test for how strong and complete the access to conventional and standard meanings that these mechanisms provide is, and if it really is the case that they allow speakers to go on meaning and interpreting without resorting to rationalization and mindreading.

As in the previous chapter, cases of misused expressions will still be basic in the assessment of the adequacy of the proposal. The disparity between L&S's view and the one I am endorsing here is most evident with respect to these occurrences. In privileging conventions above all rest, L&S's claim is that in cases of unintended meaning, as in all others, the speaker is still committed with the regular meaning of the words uttered and, in what concerns linguistic communication, with nothing more. I disagree in two respects as I will explain next.

8.2. MALAPROPISMS AND FAULTY UTTERANCES

I want now to focus on a particular type of (mis)use of language. I will be using "faulty" in application both to utterances and speakers only in cases of localized but extreme incompetence in the language of the community. Most of the cases of malapropisms noted before fit into this category - to avoid unnecessary complications, I leave out cases of intentional misuses such as those attributed to Goodman Ace in the previous chapter. I have in mind only cases of *total* misunderstanding, on the part of the speaker, of the standard meaning of the words she is using.

I expect the distinction between total and partial misunderstanding to be intuitive enough. We could perhaps make it a principled one by reference to the extension of previous, or potential, *identical* uses of the expression that the speaker would be ready to renounce were she to learn about her mistake. In a merely *partial* misunderstanding, the speaker would still hold on to the adequacy of some *identical* uses of the expression, that is, uses following the same *partially* inaccurate rule - consider, for instance, Burge's famous arthritis case that

we will discuss below. In cases of *total* misunderstanding, the speaker would realize that all such uses must be rejected, as in the previous example of "*epitaphs*" instead of "*epithets*", or in the next one of "*jeopardize*" instead of "*deputize*".

L&S put forward the following example of a malapropism they borrow from Zwicky, (1979, 341, Ex 9), where a steward utters (199)[163]:

(199) I jeopardize you to handle my duties. (Lepore and Stone, 2015, 218)

They explain that it is clear to everyone that the steward "expects" and intends to *deputize* her hearer, i.e. "to empower the hearer to act in [her] place" (ibid., 218). However, the sentence she chooses is standardly used to express a speaker's intention to *jeopardize* her hearer, to put the hearer in danger. Now, the question is: What meaning, or meanings, should be attributed in cases such as these, in cases of faulty utterances? Here is L&S's answer to this question which I quote at length since this is a very important passage in the argument:

> … our view articulates a clear standard that privileges conventional meaning in each of these cases. We look to the speaker's basic intention to settle which utterance was used and how it is grounded in the causal and informational connections to the environment that constitute meaning. Even when a speaker has false beliefs about the meanings of her words or the identities of objects in her environment, we must still characterize her contribution to the conversation by interpreting her basic intentions against the operative context. When a speaker represents the meaning of a word via deference to some network in a community, then what that word means, for the speaker, is what it means in the community. It doesn't matter if the speaker couldn't articulate what the meaning is or is committed to use some other meaning because of false beliefs she has. ... Once the speaker commits to the grammar of her utterance, she has established a basic intention that links that utterance to its meaning. After that, it's not up for grabs what the utterance means. (ibid., 219)

L&S, in precise accordance with the principles of their Direct Intentionalism explained above, defend that there is only one attributable meaning in these types of cases - the standard or community meaning of the uttered expression: in this case, *jeopardize*, and not *deputize.*

The standard and community meaning of an expression often coincides with the conventional meaning, in the Lewisian sense of *convention*, but not in cases of faulty utterances, as I will defend next. The most plausible interpretive option is to take L&S to be

[163] I am preserving L&S's numbering in order to facilitate transit from one text to the other.

using "conventional meaning" not in the Lewisian sense but simply as a loose synonym of community or standard meaning.

L&S are explicit about what, in their view, explains and justifies the standard attribution. The speaker's basic intention determines which linguistic form was actually uttered and the "operative context", the "environment that constitutes meaning", ensures that the utterance is bestowed with the expression's standard meaning for the relevant community.

They further add, and then reiterate, that, for the purpose of semantic interpretation, it matters not whether the speaker is minimally aware of the standard meaning of the expression. What is here at stake, as anticipated above, is Direct Intentionalism's tolerance for a peculiar form of *mental disintegration*. That is, L&S allow basic intentions to disagree with what else is going on at deeper levels of the speaker's mind. As they themselves put it, "Direct Intentionalism lets us characterize a speaker as meaning something that she didn't think she intended" (ibid., 219). The indexical nature of basic intentions' contents is clearly expected to play a role here. It is because speakers are not required to *conceptualize* or *descriptively represent* those meanings that they are apt to bear conflicting intentions without realizing it. As indicated above, L&S hold that "the objects of actions in basic intentions must be represented indexically; like expressions in natural language such as this and that, they refer not through a name or description but through the interaction that the agent has with them…" (ibid., 209). In this case, the object is some specific meaning that the speaker is supposedly *indexically representing* simply by picking up the correspondent word. The relevant *environment* is supposed to establish the possibility of such *interaction*[164].

At the same time, L&S recognize both the speaker's intention in action to mean something different with her words and the audience's recognition of this intention. Direct Intentionalism, they explain, does not require us to pretend that those states and understanding do not take place. L&S assure us that "[w]hen we need to, we can always

[164] L&S propose an analogy with a practical activity running against the agent's expectations and further intentions - when a cook confuses sugar with salt. The cook still forms the basic intention to add salt to the mix by employing some indexical representation of the salt that is somehow made possible in virtue of the features of the physical environment: "When a cook mixes two cups of salt into the cake batter, an enlightened observer can recognize that she must have confused the salt with the sugar. Her basic intention is clear: To mix in that stuff. Her larger intention in action is transparently predicated on the false belief that that stuff is sugar." (ibid., 219). Analogously, according to L&S's story, a speaker can form the basic intention to mean jeopardize by means of some indexical representation of that meaning that is somehow made possible in virtue of the features of the linguistic environment: "Just so for the speaker of (199), for example. His basic intention is to say I jeopardize you, meaning what jeopardize means. H[er] larger intention in action is transparently predicated on the false belief that jeopardize means deputize." (ibid., 220)

describe the speaker's intentions precisely. And we can always appeal, when necessary, to an enlightened audience's ability to recognize them" (ibid., 219). It is just that they do not take such states and understanding to be sufficient to determine proper cases of meaning.

I disagree with L&S's stance on the meaning of faulty utterances on two scores. I will deal with each question in turn. First, I will argue that in this type of case there appears to be not enough in place to commit the speaker and her utterance with the standard meaning of the expressions in question. Second, I will defend that the actually intended (in action) meaning of such utterances, especially in those cases where that meaning is adequately recognized and handled by the audience, must be reckoned as a genuine instance of linguistic meaning of some sort.

8.2.1. CHALLENGING STANDARD MEANINGS

In this section, I question the *environment*'s capacity to bind speakers to the standard meanings of expressions in cases of faulty utterances. I consider the potential of the two social structures established above as genuinely explanatory of the publicness of meanings: conventions and externalist institutions.

8.2.1.1. CONVENTIONS AND FAULTY UTTERANCES

L&S's first recommendation, as quoted above, is to *privilege the conventional meaning.* However, in a Lewisian sense of convention, there is no conventional meaning to refer to in cases of this nature. I will explain.

Consider, again, the utterance of (199). There is no relevant convention in force for utterer and hearer in this episode. We can suppose the hearer to actually participate in *the standard convention* regarding the term, the convention of using *'jeopardize'* to mean *jeopardize*. Following the type of analysis presented in the previous chapter we could define this convention as follows:

> A regularity *R*, in action or in action and belief, is a convention, in a population P, of using *'jeopardize'* to mean *jeopardize* if and only if,
>
> (1) The members of P conform to the regularity of taking *'jeopardize'* to mean

jeopardize. That is, if someone utters '*jeopardize*' that person means *jeopardize*; if someone hears it uttered, that hearer responds by taking the speaker to have meant *jeopardize.*

(2) Everyone believes that the others conform to the regularity of using *'jeopardize'* to mean *jeopardize.*

(3) This belief that the others conform to the regularity of using *'jeopardize'* to mean *jeopardize* gives everyone a good and decisive reason to conform to it.

(4) There is a general preference for general conformity to the regularity of using *'jeopardize'* to mean *jeopardize.*

(5) The regularity of using *'jeopardize'* to mean *jeopardize* has alternatives.

(6) All these facts are common knowledge in P.

Now, the steward, on the other hand, believes and acts as if both she and her hearer participated in a different convention[165], the convention using *'jeopardize'* to mean *deputize.* The steward does not fulfill the set of conditions for counting as a member of the population where the standard convention regarding the use of *'jeopardize'* is in force. She fails almost every one of them, she does not conform to the regularity, she does not have the relevant beliefs, preferences or common knowledge. In such a case I see no alternative but to conclude that the steward does not participate in the standard convention regarding the use of *'jeopardize'*. That convention is simply not in force for the steward; she is no member of P.

There are a couple of possible counter responses to this claim that are worth considering, although, I must anticipate, I will conclude against their ability to justify any retraction.

8.2.1.1.1. UNSUCCESSFUL REJOINDERS

First, there is the question of the perhaps exaggerated stringency of the conditions. Lewis himself admits that the six general necessary and sufficient conditions for there to be a convention are too strict. Instead of going for full satisfaction he says, in a brief note in his

---

[165] If, as I am assuming, this is not simply a slip of the tongue, and there is some consistency in the way the steward is using the term.

Languages and Language, that it should be enough that "they almost hold. A few exceptions to the "everyone"s can be tolerated" (Lewis, 1975, 165). In this text, however, Lewis is short of more particulars and explanations regarding more exactly where, how and why. Some further elucidation must be found in *Convention* (1969), in particular in a section named "*Degrees of Convention"*. There Lewis details a "reasonable way of relaxing" (Lewis, 1969, 76) the universal quantifications by allowing some of the participants to fail to satisfy some of the conditions, at least in some of the relevant situations. The important point is that he refuses any relaxing with regard to the knowledge conditions:

> The common-knowledge requirement involves universal quantifications over *P*... . We need not allow any exception to these; anyone who might be called an exception might better be excluded from P. (Lewis, 1969, 76-7)

Lewis is quite emphatic here. Without the relevant knowledge, he defends, they "might better be excluded" from the convention's population. See also what he has to say, a few pages earlier in the text, concerning the decisive importance of possessing the adequate (common) knowledge of the convention.

> So if a convention, in particular, holds as an item of common knowledge, then to belong to the population in which that convention holds - to be party to it - is to know, in some sense, that it holds. If a regularity R is a convention in population P, then it must be true, and common knowledge in *P,* that *R* satisfies the defining conditions for a convention. If it is common knowledge that *R* satisfies them, then everyone in *P* has reason to believe that it is true, and common knowledge in P, that R satisfies them; which is to say that everyone in P must have reason to believe that R is a convention.
>
> This is not to say that a party to the convention has any special, infallible way of acquiring his knowledge.[166] But he must *have acquired* it somehow, in an ordinary way, in order to be one of those among whom the convention holds. Discovery of the convention is the principal part of one's initiation into it. (ibid., 61-2)

As I see it, this is as it should be. No one can be said to participate in a convention without possessing the relevant knowledge - in some form or another: propositionally or not, explicitly or not[167]. That is the whole point of conventions being *rational* solutions to

[166] Of course. It is the other way around!
[167] On this issue, see, for instance, Lewis' reply to an objection*:* "Objection: If there are conventions of language, those who are party to them should know what they are. Yet no one can fully describe the conventions of language to which he is supposedly a party."

coordination problems. The agent must have a reason to conform to the regularity in question, and possessing the right knowledge is essential for that. If the agent is ignorant or too mistaken about the regularity in question, she has no reason to conform or expect conformity to it, and, naturally, she will not.

These last considerations hinge on what we above referred to as the *built-in Griceanism* of Lewis' notion. A meaning convention effectively amounts to the consolidation of the sort of collaborative interchange and meeting of minds on the meaning of some utterance that Grice identified. It engages the same type of mechanisms and mental states. A meaning convention is useful only inasmuch as it allows communicative intentions of the Gricean type to be formed and adequately recognized. Still within the Gricean model, this possibility must again be grounded in some form of propitious familiarity among interlocutors - in this case, full-fledged common knowledge. The novelty is the addition of an element of stabilization, across time and across agents.

The second possible rejoinder turns on the idea of retreating from local, *expression-level,* meaning conventions to broader, *whole-language* meaning conventions when deciding on some speaker's inclusion in the relevant linguistic population. Someone might be tempted to defend that the standard meaning is the conventional meaning even in cases of faulty utterances by holding that even when there is error or ignorance concerning particular expressions, a speaker should still be counted as a subscriber to the *larger* convention concerning the *whole* of the language to which the expression belongs. Such a respondent would likely argue that the alternative - i.e. to insist on conventional agreement at expression-level - would seem to commit us to idiolects and, perhaps, even to rejecting there actually being any shared languages, since it is more than plausible that hardly any two people share a language in the sense of scrupulously sharing all the subsidiary conventions concerning each expression of the language. On top of that, it might even be claimed to go against the spirit of Lewis' proposal. Is he not talking about conventions governing the use of prototypical languages, commonsensically individuated - English, French, Portuguese, ... - and shared by a population of more than just a few?

Starting by this last point, I find Lewis somewhat equivocal or, most likely, simply

Reply: He may nevertheless know what they are. It is enough to be able to recognize conformity and non-conformity to his convention, and to be able to try to conform to it. We know ever so many things we cannot put into words." (Lewis, 1975, 181)

neutral with respect to the dimension of the languages and populations that his proposals are meant to apply to. Here is, for instance, something he says in advocating the changes introduced in his last proposal:

> ... one advantage of the change: suppose there is only one speaker of an idiolect, but several hearers who can understand him. Shouldn't he and his hearers comprise a population that uses his idiolect? (Lewis, 1975, 170)

The implied answer is yes, but the important point about this passage is that it shows us that Lewis is not shy about the possibility of his proposed analysis individuating linguistic populations very, very narrowly.

Second, independently of Lewis' pronouncements on the topic, there really seems to be little or no point in invoking a language-level convention so as to be able to count the faulty speaker as some sort of honorary participant in the expression-level convention of which she has no knowledge. The grand idea about Lewis' account of language use as a conventional activity was that of representing episodes of communication as instances of successful coordination, supported by common knowledge and interest. That is not what happens in the type of cases signaled. If, in such cases, interlocutors flagrantly fail to share the narrower, expression-level convention, I do not see what an appeal to an *imperfect* participation in the broader one can do to help. If we were to take both interlocutors as subscribers to the same convention, broad or narrow, we should expect them to coordinate. Instead, in these cases, they do not. What is there to explain by virtue of an appeal to conventions - in the Lewisian sense of convention; of course, there can be others - is perfectly explainable taking only into account speakers' subscriptions to expression-level conventions. What cannot be explained by appeal to expression-level conventions - namely, cases of non-standard uses of expressions - can neither be explained by broader ones.

I still want to stress that, as long as the speaker is not completely ignorant of the standard meaning, there might be something to be gained from still counting her as a subscriber to the standard convention. We will reflect upon these intermediary cases next. But the matter here is different. I am talking about cases such as the one provided above, where the steward is fully and not just partially mistaken about the meaning of "jeopardize".

There may be other understandings of convention that are able to accommodate L&S's option of taking the speaker to be conventionally bound by a meaning convention that

she is unaware of. What I claim is its incompatibility with Lewis' conception. One is free to go either way. One can either defend that Lewis' conventions adequately grasp speakers' relation to their languages or admit that they ultimately individuate idiolects. I do not find this corollary that problematic, nor should it prevent us from also resorting to less rigorous notions of *public language* to talk, for instance, about Portuguese or French. Furthermore, such watered-down notions could, plausibly, be analyzable either in terms of sufficient overlapping between idiolects, or of possession, by each speaker, of a large enough fragment of some standard or official language. Or one can choose to preserve more broadly individuated languages - and, say, pick them from some official list of languages - but reject Lewis' notion as being the decisive one in linking speakers to their languages. What seems impossible is to conciliate both – Lewis' insight about speakers' relation to their languages, and broadly individuated languages and linguistic populations. In particular, nothing from Lewis' original insight seems to be of any worth in understanding speakers' relation to expressions that they radically misunderstand, as in the cases of faulty utterances. No indulgence with the watered-down notions of *shared language* mentioned should go as far as to mask this fact.

In conclusion, conventions are not doing it for L&S. No meaning convention of the Lewisian type can sanction the standard interpretation of a faulty utterance like (199). I turn now to the second social structure in the list - Externalism's institutions.

8.2.1.2. EXTERNALISM'S INSTITUTIONS AND FAULTY UTTERANCES

The question in this section is whether externalist institutions are apt to form an environment that commits speakers to the standard public meaning of the expressions they misuse in the particular cases of faulty utterances. I am not disputing externalism's general positions, as briefly explained above. It indeed appears to be the case that there are social structures, external to the individual speaker - such as mechanisms of deference and long chains of name transmission from speaker to speaker - doing very significant work in keeping terms fixed to their community meanings. These mechanisms allow the speaker's less than perfect knowledge of the standard meaning of some expression still to be compatible with her ability to use the expression to mean that standard meaning. But how far can we take this insight?

The cases targeted in the classic defenses of Externalism are those of mere *partial* incompetence with some community term. For instance, the speaker would not be able to tell the referent of “water” apart from a deceiving ersatz like twin-water (Putnam, 1975), or he would not know that “arthritis” can only be predicated on conditions of the joints, not, for instance, on conditions of the thighs (Burge, 1979). These cases are importantly different from those of faulty utterances, where the incompetence must be *total*, or virtually so.

Let us have a look, now, at what three authors from the Externalism’s canon - two of which L&S explicitly appeal to - have to say that may bear more particularly on the specific cases of faulty utterances.

8.2.1.2.1. KRIPKE

Kripke and Putnam - to whom we will turn next - hardly address cases where the speaker has wrong beliefs concerning the meaning of her words. The closest I could find in Kripke’s work is when (Kripke, 1977) he tackles the example, also invoked by L&S (2015, 220), where a speaker misidentifies a person he sees in the distance and, having that visible person in mind, utters (198).

(198) Jones is raking the leaves today.

This is no malapropism, nor even a case where the error concerns the speaker’s knowledge of the actual reference of a certain name. However, it is still a case where there is a clash between the speaker’s communicative intentions behind the utterance and the standard referent of the name he uses. The speaker intends to refer to the person he is looking at but that person is not the standard referent of “Jones”.

Again, according to L&S, the only thing here worth counting as meaning or reference in this situation is the standard meaning and reference of the words the speaker uses. In this they diverge from Kripke. They do so acknowledgingly:

> .. following Kripke (1978) among others, we can appeal to the fact that one meaning, but not another, is conventionally encoded: we can talk about what the sentence means as opposed to what the speaker meant. Moreover, details of the context may allow us to infer that some aspect of the speaker’s intention is particularly important, so that the speaker’s commitments are most faithfully laid out by backgrounding or even factoring out one or the other interpretation of his utterance. (...)

> By contrast, our view articulates a clear standard that privileges conventional meaning in each of these cases.

As they see it, the utterance meaning simply is the product of the standard meanings of each contributive expression, and "[t]here are no special meanings, over and above" (ibid., 199) that. Thus, the speaker in (198) should actually be taken to be talking about Jones, not about the person that he is looking at, not about the person he intends (in action) to be talking about.

As for Kripke, he chooses to distinguish two meanings. On the one hand, there is what the words or sentence mean - he calls it the *semantic* meaning or reference. On the other, there is the *speaker* meaning and reference, what the speaker meant and referred to with his utterance.

> Two people see Smith in the distance and mistake him for Jones. They have a brief colloquy: "What is Jones doing?" "Raking the leaves." "Jones," in the common language of both, is a name of Jones; it never names Smith. Yet, in some sense, on this occasion, clearly both participants in the dialogue have referred to Smith, and the second participant has said something true about the man he referred to if and only if Smith was raking the leaves.
>
> (...)
>
> Jones, the man named by the name, is the semantic referent. Smith is the speaker's referent, the correct answer to the question, "To whom were you referring?" (Kripke, 1977, 111-2)

I must again stress that this situation is importantly different from that of a faulty utterance. Here, the speaker's error is located not in her linguistic beliefs but in her world beliefs. However, while it is surely not enough to elucidate what Kripke's thoughts on the interpretation of a faulty utterance might be, at least it shows us his commitment to consider the speaker's intended (in action) meaning or reference when counting the various forms of meaning involved in an episode, even when that meaning is somehow in conflict with the standard meaning of the expressions used.

8.2.1.2.2. PUTNAM

In Putnam (1975), again there is little thought concerning cases where speakers bear mistaken beliefs about the meaning of their utterances. However, we find a few paragraphs concerning minimal conditions for counting as a competent user of some expression - and, so

I infer, for counting as being able to mean its standard meaning in uttering it.

> Suppose a speaker knows that ‘tiger' has a set of physical objects as its extension, but no more. If he possesses normal linguistic competence in other respects, then he could use 'tiger' in *some* sentences: for example. 'tigers have mass'. 'tigers take up space', 'give me a tiger', 'is that a tiger?', etc. Moreover, the *socially determined* extension of 'tiger' in these sentences would be the standard one. i.e. the set of tigers. Yet we would not count such a speaker as 'knowing the meaning' of the word *tiger.* …
>
> … We shall speak of someone as having *acquired* the word' tiger' if he is able to use it in such a way that (I) his use passes muster (i.e. people don't say of him such things as 'he doesn't know what a tiger *is',* 'he doesn't know the meaning of the word" tiger .... etc.); and (2) his total way of being situated in the world and in his linguistic community is such that the socially determined extension of the word 'tiger' in his idiolect is the set of tigers. (Putnam, 1975. 247)

The first condition listed by Putnam in this passage for possessing a word is that of revealing some minimal competence in its use. Other external factors can and do contribute to determining the meaning of its utterance in a sentence, but only conditionally on the satisfaction of this basic requirement. Putnam insists on this point, elaborating on examples of flagrant incompetence.

> Suppose our hypothetical speaker points to a snowball and asks, 'is that a tiger?'. Clearly there isn't much point in talking tigers with *him.* Significant communication requires that people know something of what they are talking about. ... What I contend is that speakers are *required* to know something about (stereotypical) tigers in order to count as having acquired the word 'tiger'; something about elm trees (or anyway, about the stereotype thereof) to count as having acquired the word 'elm'; etc. (Putnam, 1975. 248)

Of course, this requirement of minimal competence is not satisfied by our faulty speakers with respect to some of the words they use in their faulty utterances. If this is right, if faulty speakers are not to count as competent users of the expressions in question, on what basis are we to ascribe standard meanings to such utterances?

I believe that these passages clearly reveal that Putnam does not side with Direct Intentionalism in this particular. His brand of Externalism does not sanction standard interpretations of faulty utterances. However important the role of social mechanisms in meaning determination, Putnam does not see them to be so strong and effective as to so thoroughly dispense the speaker from contributing with the right beliefs and intentions.

8.2.1.2.3. BURGE

Lastly, there is Tyler Burge. L&S do not mention him[168] but, at first glance, and despite his emphasis on the mind, Burge's account of the famous *arthritis story* (Burge, 1979) could seem particularly congenial to L&S's standard-privileging approach. His thought experiment is meant to prompt intuitions supporting the ascription of standard meaning (and content) to a speaker's words (and mental states), despite flaws in her knowledge. Not only can the speaker mean something she does not fully know she means, the facts that determine whether she means it or not are not facts about the individual but facts about her social context.

Unlike Kripke and Putnam, Burge actually focuses on a case of mistaken belief concerning the standard meaning of some uttered expression. He tells us about a certain person, call him T, who believes he has arthritis in his thigh. He does not know that arthritis is a condition of the joints only - this is the relevant flaw in his knowledge on the topic. When he complains to his doctor by saying "I fear I have also developed arthritis in my thigh", we would still be willing to ascribe to his utterance - in particular, to his use of "arthritis" - the expression's standard meaning. That is, we would not feel tempted to endeavor any sort of reinterpretation - in particular, that of taking him to mean *tharthritis* by "arthritis", instead of plain standard arthritis, where *tharthritis* would be a medical condition very similar to arthritis except for the fact that it is not restricted to the joints.

Burge then imagines a counterfactual situation where T has exactly the same internal state and history but in which the linguistic community he belongs to actually uses the word "arthritis" to standardly mean *tharthritis*. In such a scenario, although we would still be dealing with the exact same individual, with the exact same intrinsic properties, he would be meaning a different meaning with the same utterance, and believing a different belief, specified in terms of *tharthritis* and not arthritis.

There is no call for reinterpretation in this story. However, the important thing to hold firmly in mind here is that the arthritis case is still not the case of a real faulty speaker and utterance. We are dealing here with partial, not total, misunderstanding. What is more, Burge

[168] There is this sweeping reference to externalism. "Our particular proposal targets an externalist view of meaning, grounded through a division of linguistic labor, but still emphasizes the role for coordination in what's happening in a conversation, moment by moment." (Lepore and Stone, 2015, 255-6)

actually addresses these latter types of cases. In such cases, he argues, reinterpretation might actually be called for.

> There are, of course, numerous situations in which we normally reinterpret or discount a person's words in deciding what he thinks. Philosophers often invoke such cases to bolster their animus against such attributions as the ones we made to our subjects: 'If a foreigner were to mouth the words "arthritis may occur in the thigh" or "my father had arthritis", not understanding what he uttered in the slightest, we would not say that he believed that arthritis may occur in the thigh, or that his father had arthritis. So why should we impute the belief to the patient?' Why, indeed? Or rather, why do we? (118)
>
> Malapropisms form a more complex class of examples. I shall not try to map it in detail. But in a fairly broad range of cases, we reinterpret a person's words at least in attributing mental content. If Archie says, 'Lead the way and we will precede', we routinely reinterpret the words in describing his expectations. Many of these cases seem to depend on the presumption that there are simple, superficial (for example, phonological) interference or exchange mechanisms that account for the linguistic deviance. (Burge, 1979, 119)

In the same text, even after stating his view - manifestly close to that of L&S - that "wherever the subject has attained a certain competence in large relevant parts of his language and has (implicitly) assumed a certain general commitment or responsibility to the communal conventions governing the language's symbols, the expressions the subject uses take on a certain inertia in determining attributions of mental content to him" (ibid., 147), Burge is quick and explicit in qualifying it, that is, in recognizing exceptions and the occasional need for reinterpretation.

Once more, two pages ahead, and close to the end of the article, he affirms the complex nature of the object, and reiterates the disruptive potential of faulty utterances. He lists "[i]nsincerity, tongue slips, certain malapropisms, subconscious blocks, mental instability all make the picture more complex" (ibid., 149), and recognizes that these, among other things "[affect] the inertial force of 'face value' construal". This he still deems compatible with "keep[ing] steadily in mind the philosophically neglected fact about social practice: Our attributions do not require that the subject always correctly or fully understand the content of his attitudes." (149-50)

Finally, Burge even returns to the topic in his 2006 postscript to the article to reaffirm his first positions:

> … there are many situations in which reconstrual is appropriate. But neither reconstrual nor

> ‘homophonic’ interpretation is automatically correct. Reconstrual is correct in many fewer instances than the common philosophical wisdom maintained three decades ago.
>
> Reinterpretation picks up on something in ordinary practice. Misuses or failures of understanding exemplified by malapropisms, tongue slips, extreme ‘category’ misuses, the first uses of words by very young children, and the fumblings of foreigners, all normally and rightly occasion reinterpretation. Most other cases are more complex. (Burge, 2006, 176)

I follow Burge both in his externalist, or, as he prefers, *anti-individualistic*, general point - that there is ordinarily more to one’s meanings and contents than one is aware of - and in his attention to the complexity of the distinction between cases where standard meaning ascriptions are due and cases where they are not. Although declining the attempt to draw a sharp line[169], he is perfectly mindful of the fact that different cases can perfectly well justify either one of those options.

L&S reject this distinction by refusing one of its terms - reinterpretation. Apparently, to Direct Intentionalism the only question is whether or not the speaker counts as belonging to a certain linguistic community. This is what determines the confines within which one is justified in ascribing standard meanings and thus, of course, invites the question: what determines whether or not the individual belongs to some linguistic community?

The two questions - to reinterpret or not to reinterpret, and to count as a member of some linguistic community or not - are obviously related but there can be different takes on what exactly that relationship consists of. Adequate face-value interpretation and community membership in an intuitive or pre-theoretic sense tend to coincide, but can also fail to. The adequacy of face-value interpretation is not automatically warranted for members of some linguistic community loosely delimited. Burge clearly expresses this view in the quoted passages. My resistance to standard ascriptions to faulty utterances does not amount to, or depend on, the rejection of a broad and commonsensical individuation of linguistic communities - as long as, with Burge, one preserves the possibility of occasionally, when justified, refraining from ascribing to some utterance its standard community interpretation.

8.2.1.2.4. EXTERNALISM ON SEMANTIC AND PSYCHOLOGICAL INTEGRATION

[169] “Neither reconstrual nor standard construal is automatic. The relevant conditions governing each are extremely complex and varied.” (Burge, 2006, 176)

A line must be drawn somewhere, but the question is: Where and on what reasons? An involuntary sound is clearly not a candidate for a standard interpretation, but perfect control of the meaning of one's words is also too much to ask. In between we find faulty utterances but also smaller faults, as in the arthritis story. I defend the view that the steward's utterance (199), along with other *faulty* utterances, belongs outside the divide where standard meaning ascriptions are due, but I follow the orthodoxy in finding no need to reinterpret the confused patient's utterance about his thigh problem.

The fact that there is a distinction to be made, one that might not be so obvious or automatic as one might have expected, becomes perhaps more vivid in contexts of language acquisition. Next, I describe some typical episodes, arranged in two different sets, that illustrate different stages in a path that might be described as leading from *fully not meaning* some expression's standard meaning to *fully meaning* it.

Think, first, of some random infant babbling that generates some sound resembling some word. Now move from this to the case where the infant is still learning her first language and includes in her speech some complicated word - for instance, "mortgage" - she does not understand at all, and has just picked up from the radio. In a different type of case, the child's partial competence with some term is undeniable - say, for instance, with "horse" - even if there are still some important adjustments and fine-tuning to be accomplished - imagine, for instance, that she is using the term to refer also to donkeys. Finally, consider some (probably older) child's close to flawless mastering of some term, perhaps, "mother".

For the second series of cases, start by thinking of someone who has produced some meaningful utterance in her own language whose sound is indiscernible from that of a different sentence belonging to a different language. Now imagine the case of a confused tourist in a foreign country painfully trying to articulate the wrong sentence. Imagine that she has extracted it from some coursebook but copied the wrong sentence into her notebook; she wanted to *ask for the menu* but instead articulated the sentence that is standardly used to *ask for the bill*. Thirdly, consider the case where the person partially grasps the meaning of some foreign word from watching natives using it, but extends its application beyond the word's proper borders. Take, for instance, the name "pastel de nata", which refers to a specific Portuguese pastry, a small puff pastry tart filled with a sweet cream made from milk and eggs. After a small number of observations, the person learning the term might extend it, incorrectly, to a second type of pastry, one with the exact same shape and filling but with a

shortcrust pastry cup instead of a puff pastry one - this second pastry usually goes by the name “bom bocado”. Finally, there are the regular cases in which the tourist has effectively acquired (close to) perfect competence with some common term, maybe one to which there is some easy and clear-cut translation in his native language.

The same questions from before are pertinent here for both groups of examples. Which cases are due a standard interpretation and which are not, and following what reasons or principles? Where to draw the line and why?

I do not think one should hope for a general and precise method of operating the relevant distinction, but this should not be taken to imply that nothing substantial can be added on the topic. In this section, I return to a popular and recommendable idea of semantic and psychological integration - one that, in particular, is integral to Davidson’s Principle of Charity - that goes a long way to determining a principled and somewhat intuitively appealing distinction among the type of cases listed so far. Importantly, this idea is perfectly compatible with Externalism, and, furthermore, it is even endorsed by the leading externalists we have been considering.

Burge, Putnam and Kripke, along with many others, affirm a natural solidarity and continuity between our mental states and our language use. There are essential links between assertions and beliefs, questions and uncertainties, requests and desires, and more. Accordingly, the interpreter should only ascribe the meaning if he is also willing to ascribe the correspondent mental state. This idea could perhaps be developed into more precise principles or norms of interpretation such as:

a. The interpreter should not take a speaker *a* to be honestly asserting *that p* with her utterance of some expression *s* if he is not also willing to ascribe to *a* the belief *that p*, (even if *s* is standardly used by speakers in the relevant vicinity of *a* to assert *that p*).

These considerations of semantic and psychological integration appear powerful enough to determine a clear separation of the cases so far considered. In particular, the speaker’s uses of “horse” and “pastel de nata”, together with that of “arthritis”, appear to go together with thoughts involving contents that correspond to the standard meanings of those expressions. It is likely that the interpreter would still be taking the child to be thinking of horses, perhaps something like: *she thinks the donkey is a horse*; the foreigner to be thinking

of *pasteis de nata,* perhaps: *she has mistaken a bom-bocado for a pastel de nata*; just like he took the patient to be thinking of arthritis: *he believes he has arthritis in his thigh*. That is the natural thing to do, and arguably the best way to describe and make sense of their words. In striking contrast, the interpreter would not be willing to ascribe to the child any thought about mortgages and he would not link the foreigner's utterance with any thought regarding the bill, just as he would not associate the steward's utterance with any wish or intention of putting her hearer in jeopardy.

As pointed out, externalists endorse this conception of integration of domains. L&S, on the other hand, seem bound to the possibility of there being exceptions to principles or norms such as *a.* above, and, in general, take a less strict stance on the conditions and requisites of solidarity and continuity between our thoughts and meanings. Let us compare them more carefully.

Take Putnam first. Although the special focus in *The Meaning of 'Meaning'* is on language, afterwards Putnam comes to notice that the same points must hold for the mind so as to ensure the right type of continuity and solidarity between propositional attitudes and meanings. See, for instance, what he says in his '*Why Functionalism didn't work?'*.

> The first difficulty I encountered with my functionalist views was that they were incompatible with the account of meaning that I myself put forward in 'The Meaning of "Meaning"'. According to the arguments of that essay, the content of our beliefs and desires is not determined by individualistic properties of the speaker but depends on social facts about the linguistic community of which the speaker is a member and on facts about the physical environment of the speaker and of the linguistic community. (Putnam 1992, p. 443)

When the speaker's utterance - '*Elms are deciduous trees'* - is about elms and not about beeches, despite the speaker's inability to tell one kind from the other, so is his correspondent belief, i.e. the belief *that elms are deciduous trees.* The reference/meaning ascription goes together with an identical content ascription to the speaker's mental contents. Or, in Putnam's words from that same essay, "the interpretation of someone's language must always proceed simultaneously with the ascription of beliefs and desires to the person being interpreted." (Putnam 1992, p. 446)

The same continuity is even more vivid in Burge, for whom, from the start, the prime concern was the ascription of mental states, not of meanings. That the inference from meanings to mental attitudes must be valid is the tacit assumption at work throughout the

whole essay. Here, for instance, it emerges explicitly:

> … wherever the subject has attained a certain competence in large relevant parts of his language and has (implicitly) assumed a certain general commitment or responsibility to the communal conventions governing the language's symbols, the expressions the subject uses take on a certain inertia in determining attributions of mental content to him. In particular, the expressions the subject uses sometimes provide the content of his mental states or events even though he only partially understands, or even misunderstands, some of them. (Burge, 1979, 147)

When theorists - in particular, those of an externalist extraction - allow speakers an incomplete understanding of their utterances' meanings, they naturally tend to extend the liberality to thinkers' mental attitudes, allowing them to incompletely grasp the notions present in their thoughts. Manifestly, Burge's view sounds a lot like that of L&S - notice, in particular, the last sentence from the passage. However, there is the apparently minor detail of how to deal with extreme cases such as malapropisms to make all the difference. The need to reinterpret in such cases - defended by Burge and rejected by L&S - forces the introduction of a small amount of mindreading and rationalization into the process of interpretation - an initial small amount which, I believe, we can capitalize on.

Finally, there is Kripke. The commitment to the kind of agreement between meanings and mental attitudes that I want to show to be shared even among *externalists* could, just as easily, have gone unarticulated by these authors. However, again by a happy coincidence, at a certain point Kripke also happens to state this view very clearly:

> Let us make explicit the disquotational principle presupposed here, connecting sincere assent and belief. It can be stated as follows, where 'p' is to be replaced, inside and outside all quotation marks, by any appropriate standard English sentence: "If a normal English speaker, on reflection, sincerely assents to 'p', then he believes that p." (...)
>
> Taken in its obvious intent, (...) the principle appears to be a self-evident truth. (Kripke 1979, 137-8)

Of course, this essay, *A Puzzle about Belief*, is all about a puzzle that can be derived from accepting this principle together with little more. Despite the puzzle, the principle retains its cogency. One would be hard pressed to justify its abandonment.

This is the Externalist standard. To some degree, our three canonical authors elaborate on the idea of continuity between meanings and thought. The three hold that, for a speaker to mean something with her utterance, that meaning must be sufficiently integrated in the

speaker's mental life. Lack of details and possible discrepancies among them notwithstanding, the three are committed to something much stronger than L&S's minimal concession to psychology, namely, that to mean something with an utterance requires only basic intentions.

L&S gladly accept that broad semantic and psychological integration is ordinarily obtained. For instance, when recapitulating some of their points, they claim that:

> [w]hen a speaker utters the sentence, *Some people own cars*, she normally intends to instruct her audience that some people own cars; when she utters the sentence, *I'm happy*, she normally intends to get across to them that she is happy; when she utters the sentence, *Turn off the TV*, she normally intends to instruct her audience to turn off some particular television; and so on. Intentionalism—in the broadest sense—aligns with the intuition that meaning is the product of a deliberate action, a choice, or commitment on the part of the speaker. (Lepore and Stone, 2015, 225; see also 213)

But what is crucial here is that they take any such further integration to be unnecessary.

> … others take these considerations to establish a fundamental link between meaning in language and what we have called communicative intentions in action. But, as we have seen, direct intentionalism offers another explanation. The idea is that the [basic] intention determines what you said: What words you used, what syntactic analysis you had in mind, and so forth. You had to represent the utterance to say it, that's all. (ibid., 225)

I can easily agree with L&S in the claim that a large portion of meaning making and interpretation consists in the employment of peculiarly linguistic skills and knowledge, in a very specific type of exercise that takes place somewhat apart from the rest of one's mental life. This may very well constitute the bulk of our use of language. Where L&S and I disagree is in their readiness to accept that the connections between such specifically linguistic states and operations and our further mental goings-on, can, sometimes, be completely severed. As I understand it, this is what an ascription of standard meaning to (199) amounts to in the end.

L&S recognize that the "speaker of (199) intends [in action] to contribute that the hearer is empowered to act in [her] place" (ibid., 218). And so, presumably, that she also has some of the further typical mental states that go naturally together with this intention. Say, for instance, the desire to empower the hearer to act in her place, the belief that, through her utterance, the hearer will be empowered to act in the steward's place, together with many

others along a spreading network of states exhibiting cohesion and rationality. At the same time, the speaker clearly has no intention - no intention in action and no future intention - to use the utterance to contribute, say or mean *that the hearer is jeopardized to act in her place.*

But then L&S add another intention to the lot: the speaker's "basic intention (...) to say I jeopardize you, meaning what jeopardize means" (ibid., 220). This intention is supposed to be the one most directly responsible for her uttering of (ibid., 199) and, ultimately, the one that decides upon its meaning. Here, where the basic intention is in conflict with the intentions in action and, according to L&S, must still prevail as the determiner of utterance meaning, is where the radical severing takes place. The basic intention is peculiarly allowed to drift apart from the rest of the mental system. In particular, nothing relates it to the most relevant states of the speaker at that moment - her intention in action to mean something else, her desire to mean something else, her belief that she is meaning something else.

It is crucial not to confuse this *basic intention* with the perfectly regular *intention in action* to use the sentence to mean its standard meaning. This second intention in action is, as I see it, not fulfilled, but its place and role in the speaker's mental system are perfectly clear and raise no perplexity. This second intention in action is well integrated in the system in virtue of the speaker's false belief concerning the standard meaning of her utterance.

Contrastingly, the speaker's supposed basic intention to mean *jeopardize* finds no accommodation in the rest of the speaker's mind. As noted above, L&S appear to have managed to prevent blatant contradictions between the basic intentions and other mental states by determining that "the objects of actions in basic intentions must be represented indexically" (ibid., 209). According to Direct Intentionalism, the meaning that the speaker conveys to her utterance through a basic intention is not entertained by the speaker in his mind. Instead, the speaker is meant to grasp it through an indexical representation, the community words - as they put it, "it's as if a word points to its interpretation in the larger community" (ibid., 209). Not entertaining the content in her mind explains how she could have missed the discrepancy between, for instance, her basic intention to mean *jeopardize* and her intention in action to mean *deputize*. However, this is hardly the beginning of a persuasive account when what is being proposed is a novel category, *basic intention*, one, in particular, that is expected to ground a deep revision of a stable and intuitive doctrine. Even if a briefly explained invocation of indexicality may hint at how L&S can hope to avoid immediate troubles in their proposal, it still leaves us a long way from a satisfying elucidation

of basic intentions' peculiar nature and autonomy.

I leave the matter at this stage. In later sections I will elaborate on the advantages I find in semantic and psychological integration. Here the principal aim was to reveal the unanimity among our three externalists regarding the need or effectiveness of this type of integration, and to contrast their position with that of L&S. I believe that this common position might, to some degree, illuminate the previously identified affinities among the classic externalists, and contrasts between them and L&S. In contrast to L&S, Kripke, Putnam and Burge - as we saw in §8.2.1.2.1, §8.2.1.2.1 and §8.2.1.2.3 - share the common assumption that the speaker's psychology (beyond their basic intentions) contributes to the determination of the meaning of her utterances. The last two go even further in explicitly opposing the idea, defended by L&S, that faulty utterances call for a standard interpretation. The requisite of semantic and psychological integration emerges here as a decisive issue when one is facing the problem of how to draw and defend a principled distinction between cases where the standard or community interpretation is due and cases where it is not. The acceptance of the requisite generates a delimitation of cases that is at odds with L&S's position on the matter of faulty utterances, but agrees with what I am defending here, and with what Putnam and Burge have to declare on the topic; Kripke does not directly face the question.

Since §8.2.1.2, I have been investigating the social mechanisms and structures revealed by semantic externalists with the goal of ascertaining how extensive their role and power are in preserving the standard meaning of words in a community. The three classical proponents of Externalism that we have been following did not present these social mechanisms and structures as something that could dispense speakers from personal competence and mental engagement. They do not see these structures as completely eclipsing any need for speakers to know and intend their utterances to mean what they mean. According to their picture, externalist institutions play a very substantial part in linking words to their meanings, but there is still more to meaning something with an utterance besides being a member of a somewhat loosely defined linguistic community.

In response to the question that launched these last sections (from §8.2.1.2. onwards), I am convinced that we must conclude that externalist institutions - like conventions before them, in my argument - are not apt to constitute an *environment* that is such as to commit speakers to the standard public meaning of the expressions they thoroughly misuse, as in the

specific cases of faulty utterances.

We have now considered the two social structures put forward by L&S - only the ones established above as genuinely explanatory of the publicness of meanings: conventions and externalist institutions - and found both incapable of securing all that L&S expect from them. I conclude that L&S do not manage to show that there is enough in place to bind a faulty speaker to the standard meaning of the misused expression. I am actually convinced that there is not. I am convinced that these utterances constitute the class of linguistic uses for which it is most flagrant that automatic standard interpretation is not viable, and regarding which the interpreter has no option but to resort to mindreading and rationalization.

8.2.1.2.5. EXTERNALISM AND CHARITY

In this section I will take a brief excursus to consider and try to dispel the suggestion of incompatibility between Content Externalism and the Principle of Charity. I will address in turn three potential sources of apprehension.

First there is the danger of going too far on the externalist side. Some wariness may arise when comparing the two positions along its broadest lines. Content Externalism stresses the role of factors external to the individual thinker and speaker in the determination of her thoughts and meanings. The Principle of Charity, in contrast, commands us to focus on the properties of the individual, the ascription of meaning or thought must fit the person and her conduct. However, there is actually no conflict here, at least as long as one remains within the traditional limits of externalism. As we saw in the previous section, the three classical authors actually expect the interpreter to take more than the propitiousness of the social and physical environment into account. According to the externalists in question, the speaker or thinker must still do her share to meet some minimal conditions of competence with the word, however thin or vague these conditions might be. Here is where the Principle of Charity intervenes, ensuring that the ascription suits the person in rendering her rational. There is, thus, no evident difficulty in accommodating the Principle of Charity within an Externalist framework as long as one avoids extreme forms of content externalism, as I think one should.

Second, things can also go astray by pressing too far in the opposite direction, that is, by interpreting too strictly the demands of Charity. One runs into trouble - conflict with externalism is just one problem - if one fails to acknowledge that the Principle of Charity

leaves room for some error and disagreement (See Chapter 2 of this dissertation). To condition understanding and communication to perfect agreement among interlocutors would render it, at best, extraordinarily rare. That is not how we see or want to see things. Often we are ready to accommodate even quite significant differences. To use a typical example, despite the tremendous distance between contemporary astronomy and ancient Greek astronomy it can still be perfectly charitable to interpret them to be talking about the stars. Interlocutors should still be able to share a subject even if they do not agree on all their opinions about it - the Principle of Charity must not be taken in such a severe and implausible form that it could collide with this claim. Here is another example put forward by Putnam:

> All interpretation depends on charity, because we always have to discount at least *some* differences in belief when we interpret. For example, suppose we are reading a novel written two hundred years ago in English, and we encounter the noun "plant". In a normal context, we do not hesitate to identify this "plant" with our present English "plant"; yet, in so doing, we are ignoring a host of differences in belief. For example, *we* believe that plants contain chlorophyll, we know about photosynthesis and the carbon dioxide-oxygen cycle, and so on. These things are central to our present notion of what a plant is. All of these things were unknown two hundred years ago. (Putnam, 1992, 13)

On the other hand, of course, there are limits to Charity. For instance, one might refuse to go as far as to accept “sandwich” to refer to pizzas, even folded ones (Ludlow, 2014, 10-3), and surely a child will protest if someone uses “doll” to refer to a Spider-man action figure (Ludlow, 2014, 9-10). Putnam offers yet one more example:

> … the great metallurgist Cyril Stanley Smith once proposed to me (as a joke, but one with a serious point) that there really is such a thing as *phlogiston* (the substance that was supposed, before the role of oxygen was discovered, to account for combustion by *leaving* the burning substance and gradually saturating - or “phlogisticating” - the air). Phlogiston, Smith suggested, is *valence electrons.* What makes this a joke is that, as Smith perfectly well knows, we do not speak as he "proposed" we should; we are not prepared to say, "Phlogiston theorists were talking about valence electrons, but they had some of the properties wrong." That would be excessive “charity”. (Putnam, 1988, 13-4)

These few cases are already suggestive of the amplitude and complexity of the problem. Very often, the decision regarding whether or not to assume synonymity, whether or not we are talking about the same thing, will not be obvious or automatic. Rationality considerations will not be the only ones influencing a verdict but, I am convinced, in the broad sense in which I am using the term, they will always be decisive. The Principle of Charity is meant to guide the interpreter as he negotiates his way through the dense

environment of human communication, where different semantic-relevant facts may push in different directions, determining that ultimately he must choose whatever allows him to make best general sense of his interlocutor. The principle must not, instead, be distorted into an intransigent requirement of perfect correctness and agreement among communicators - that is the crucial point to be retained at this stage.

I encountered yet a third form in which the impression of conflict between Externalism and Principle of Charity might arise. Again the problems result from aiming too high on the side of adequacy between individual and content. The question arises at the local level when the question is how to interpret an individual who is partially mistaken about the community meaning of one of his words. Think again of Burge's arthritis story. Now the question is: would it not be more charitable to interpret the patient to be talking and thinking about *tharthritis* - a medical condition very similar to arthritis except for the fact that it can also affect the thighs - instead of arthritis? If the answer is yes, we will have a conflict, since from an externalist perspective, the right attribution would still be *arthritis*.

I will sketch two different answers to this challenge, both defending that, in these types of cases, the best way to conform to Charity and to render the speaker as rational as possible is to opt for the standard or community meaning.

The first response is to admit that, at the local level, reinterpretation allows the interpreter to represent the speaker in a better light than the standard interpretation, but to show how such an initial improvement would result in much bigger losses in the longer run. Henry Jackman (2003) develops a similar line of response, and with more detail than can be fitted into this section.

As characterized in the first part of this dissertation, Charity is a global principle, in the sense that it requires the interpreter to privilege the fitness of the whole mental system above that of any particular thought. We also saw that instead of allowing the interpreter to treat indistinctly the contribution of every thought to the general rationality of the agent, Charity expects the interpreter to be able to differentiate among states and commitments. Some flaws and revisions, given the place and role of the relevant states in the system - where one should reckon the states' centrality, importance and value for the agent, the amplitude and depth of their inferential connections, their epistemological plausibility, and more - will be much more serious than others.

The interpreter cannot decide what is the charitable interpretation by simply comparing alternatives for particular thoughts on particular occasions. He must, somehow, be able to ponder each candidate interpretation's contribution to the complex bigger picture. While a *tharthritis* ascription would save the speaker from one mistake on that occasion, it would afterwards clash in much more devastating ways with other aspects of his mind and conduct. How can we explain the patient's reaction when he comes to learn from the doctor that arthritis can only affect the joints - the fact that he will simply correct his ways instead of replying that he actually meant a different thing with the term? Or imagine, for instance, that he forms "the belief that Calcium supplements promote arthritis simply because [he hears] a doctor say 'Calcium supplements promote arthritis'" (Jackman, 2003, 155). As Jackman points out, this would "manifest [his] implicit commitment to [his] meaning the same thing by 'Calcium' and 'arthritis' as doctors do" (ibid.). This, in turn, would arguably manifest the speaker's deeper and implicit commitments and expectations regarding how language and meaning works, where externalist mechanisms and institutions are conceded a crucial role. All this would not fit well with a reinterpretation, *tharthritis* instead of *arthritis*. It seems, thus, that a thoroughly charitable assessment agrees with externalism in recommending the community meaning ascription in this type of situations.

The second response consists in noting that the same type of reasoning that would justify reinterpretation in this particular case would likely justify or require reinterpretation for an intractable number of additional cases, quickly compromising the actual possibility of interpretation.

This time, the key to reconcile Externalism and Charity is the recognition of how interpreters are limited in their resources. Interpreters can only employ their own language, thought and standards of rationality. They must do their best to fit the other into their own scheme of things, and they have limited room for change or innovation in that scheme.

Here, more concretely, is what I have in mind. Imagine that the interpreter were indeed to try to coin a new term or concept, *tharthritis,* to allow him a more faithful grasp of the speaker's ideas. The problem is that, more likely than not, the speaker will also have less than perfect understanding of the community meaning of (just to begin with): "thigh", "medical condition" and "joints". This may not have already become as flagrant for the interpreter as the speaker's incompetence with "arthritis", but it is to be expected. After all, who can boast perfect competence and knowledge of the community meaning for the

generality of our terms - in particular, if we are allowed to invoke counterfactual situations, perhaps involving twin planets, to challenge it. If this is so, would the interpreter not be also committed to the reinterpretation of the very terms he would be using in the first reinterpretation? In using “arthritis”, the patient has in mind a certain medical condition that can also occur in the thighs - that much the interpreter seems to manage to grasp; except that, plausibly, the patient is not exactly thinking of *thighs* or *medical conditions*, but of *ththighs* and *thmedical thconditions,* whatever those might be. Where would this kind of strategy lead the interpreter? How far could he extend it? Not very far. Were the interpreter to try to use new or modified concepts tailored to fit all peculiarities of his interlocutor’s mind, he would very soon be lost. A first reinterpretation undertaken in this spirit seems already to commit and tangle the interpreter in a content holism of the strictest and most inhospitable type. Obviously, the interpreter would not be able to find his way in such a maze; he would not actually be able to produce any reinterpretations of the speaker, let alone reinterpretations that would render her more rational.

Sometimes, as defended in the previous sections, reinterpretation really is the best option. This is often the case when there is available a better term or concept, as is the case with malapropisms. Perhaps, also, every once in a while it would actually be pertinent for the interpreter to try some new or modified concept. But it appears that such changes and adjustments cannot be done massively. The interpreter would lose track of any sense or meaning. The possibility of making sense of others seems to depend on not moving too far from one’s own views and conceptions, where the standards of rationality are keyed in a much more concrete fashion than the frequent emphasis on the formal aspects of rationality would lead us to expect.

The point to retain from this second exploratory response is that reinterpretation should not be too liberally attempted. Unless contained within strict limits, it undermines interpretation.

Concerning the remaining challenge - that of providing some answer to the question: “What distinguishes cases where reinterpretation is genuinely called for from the others?” - on my part, I find little to add here. Plausibly, there will be no general and precise answer to the question, and no clear and effective recipe for applying the distinction. In any case, we, everyday interpreters, are continuously faced with the problem and cannot avoid taking positions in concrete cases, however tentative and fallible. Charity, as always, will be

essential in attempting to strike some sensible equilibrium between familiarity and the inertia of standard ascriptions, on the one hand, and readjustment and accuracy in the representation on the other.

In cases like the arthritis example, reinterpretation appears not to be the best option. There seems to be enough in place to justify a standard ascription better than any available alternative ascription. Overall, given what is known of the speaker and the resources available to the interpreter to make sense of his utterance, not only on that local and circumscribed occasion but also, as we saw, in the longer run, the standard ascription fares better in rendering the speaker rational, and is hence in conformity with charity.

8.2.1.2.6. CONCLUSION ON STANDARD MEANINGS

My first aim in these last sections - §8.2.1.1 and subsections - was to show that, in the specific type of case under study, the connection of faulty speaker and her faulty utterance to the uttered expressions' standard meaning is much more tenuous than might have been expected, and L&S seem to assume. As I see it, following the reasons and explanations provided, a standard interpretation of a faulty utterance would be virtually orphan of a foundational semantics. No facts convincingly support the hypothesis that the speaker or her utterance meant what the uttered expressions are standardly used to mean in the relevant community.

In particular, I have tried to show how L&S's appeals to conventions and to the social institutions associated with the externalist tradition cannot help here. First we saw that the relevant convention is not in force for the speaker in cases of ignorance of the standard meaning. This is decisive, I argued, to render a Lewisian notion of convention useless to model any alleged commitment to the community norm.

Next, I defended that habitual forms of content externalism do not sanction a standard interpretation in such cases either. The presence of social structures and mechanisms of the type brought into center stage, some decades ago, by Kripke, Putnam and Burge are still not enough, by themselves, to invest an utterance with the standard meaning of the uttered expression. Speakers must also contribute with some minimal competence with the used terms. An adequate external environment still requires the speaker to be adequately engaged - the right mind, the right expectations and intentions - for a standard interpretation of the

utterance to be justified. I further uncovered and explained how an imperative of psychological integration articulates a good reason to hold on to this requirement. The discussion finally led us to consider the fit between externalism and the principle of charity. I have tried to dissipate some potential doubts concerning their compatibility.

There is still a tenuous link. In general, the faulty speaker still counts as a member of some vaguely delimited community of speakers of some loosely individuated language. It may also likely happen that, besides the communicative intention, the speaker holds the additional intention of conforming her use of the words to that of her community - we will consider this idea with some extra detail in the next section. In the end, I defend, this is not sufficient to justify the standard interpretation. But there is room for some nuance here. A more careful conclusion should, perhaps, take into account different senses or notions of meaning. As explained, my privileged notion is that of *first meaning* - that is what I first have in mind when I talk about interpretation. Clearly, that is not the case with L&S. This means that, to a certain extent, we are talking past each other when they affirm that the right interpretation of the utterance is the standard one, and I argue that it is not. But there is more than mere equivocation going on here. There is real disagreement, just not one that is neatly contained in the manifestly contradictory judgments. It is diffused through various levels, topics and questions: What should linguistic interpretation aim at? What should semantics be concerned with? What is prior to what? What is derived from what? What is primitive? and so on. This makes it harder to address.

Accordingly, I concede that, in some sense, the used word, “jeopardize”, means *jeopardize*. In other words, I concede that there is room for a semantic notion that captures the indisputable fact that the steward has used a word that members of the English-speaking community standardly use to mean *jeopardize.* Just as the child from the example above has used a word that, in some sense, means *mortgage*. Once again, I must insist, this - the use of a word with a standard value in the community with no awareness of what that meaning is - is importantly *less* than participating in a convention to use the word in such a way, or to be able to enjoy the advantages of the division of linguistic labor with respect to the term in question. It is very important to realize that “jeopardize” will not be able to preserve its meaning in the same way “arthritis” is able to preserve its meaning in the famous example. The word may, in some sense, preserve its standard meaning, but the speaker will still not be able to rely on social institutions to mean the meaning he completely fails to grasp.

Only in a very marginal sense and on a diminute set of occasions will a standard interpretation of a faulty utterance, one that privileges the word's ability to preserve its meaning independently of the context, be called for: occasions where the question is not one of understanding but, for instance, to assess someone's legal responsibility for her words come to mind. Such cases, as far as I can see, will have little to do with understanding, communication or coordination.

When understanding and communication are concerned, we need more than mere belonging, in some vague sense, to the linguistic community where such a usage is standard. Speakers' minds and intentions also need to be where the community standards are. In ordinary situations, we have more than just the mere standards; we also have the right intentions backing up the standard interpretation. We see this from how the standard interpretation falters when the right thoughts and intentions go missing.

The conclusion is meant to be general, even if based on reflection on a specific type of case - faulty utterances. L&S presented a case which, if accepted, would show it possible for speakers to mean their expressions' standard meanings independently of their thoughts on the matter. In arguing that this is not possible, that the strategies delineated by L&S - exploration of conventions and of externalist institutions - are still insufficient to secure the possibility of *thoughtless* meaning, I mean to restore the general validity of the claim that meanings and thoughts go hand in hand in communication. This coincidence of meanings and thought is no accident, something that could perhaps go missing in some cases. Instead, I maintain, it must be accepted as an essential condition in genuine linguistic communication.

8.2.2. DEFENDING INTENDED MEANINGS

As declared above - end of §8.2. - I disagree with L&S's assessment of faulty utterances on two scores. In the last sections - §8.2.1. and subsections - I have presented my case against their recommendation to remain faithful to standard meanings even with respect to cases of faulty utterances. It is now time to elaborate on my second challenge to Direct Intentionalism. In the following sections, I argue for a notion of intended, potentially non-conventional and non-standard, linguistic meaning - one, in particular, that would allow us to recognize the meaning actually intended by the speaker and adequately recognized and handled by the audience in such benign cases of faulty utterances as (199) as a genuine

instance of linguistic meaning of some sort.

This second challenge is directed at two twin aspects, or theses, of L&S's position - Direct Intentionalism. We can formulate the first as:

i. Genuine meaning can only be built over a conventional structure, or, more broadly put, can only be produced by means of some special social competence.

And the second as,

ii. The adequate transaction of communicative intentions is, by itself, not enough to generate meaning.

These are two sides of the same coin. In L&S's model, to press for the first is to press for the second. If conventions are needed, it is because the commerce of intentions is not enough, and vice versa. They are convinced that genuine meaning cannot rest solely on common understanding, among interlocutors, of the point behind each utterance - the meaning the speaker intends for her words. They are convinced that conventions must be added to the recipe. Mutual understanding, they sustain, must be backed up by adequate social structures.

This is how I understand L&S. I am opting for a radical reading of their conventionalism. To be fair, there is some margin for uncertainty here. I do not want to omit the fact that there are passages - both in the book and in subsequent replies to the critics - that might suggest a more moderate position on their part. For instance, they conclude a recent text in response to a group of commentators with a modest assessment of their main claim in the book. They assert that their "contention in *Imagination and Convention* is that the conventions of meaning go much further in settling the content of our utterances than our critics believe." (Lepore and Stone, 2016b, 652). Obviously, there is a huge difference between this claim that *there is more convention to meaning than traditionally assumed*, and i. above, the much stronger claim that *all meaning is conventional or involves the exploration of acquired social competences*. Despite occasional passages such as the last one, I believe the stronger attribution is still firmly justified. The stronger claim is endorsed in a vast number of passages, for instance when they characterize their enlarged notion of semantics:

> … semantics describes interlocutors' social competence in coordinating on the conversational record.
> … it's clear, because of the arbitrary nature of the record, that no matter how you think about updating the record, there's going to be a need for appropriate social competence in doing so.
> … Thus, social competence and its prototypical case, convention, becomes the key diagnostic for semantics (Lepore and Stone, 2015, 256)

When they assert that all semantic contributions require conventions or social competence - while, in their view, pragmatics "never contributes content" (ibid.83) - surely they have to be taken to be sustaining that all meaning is conventional. And there is more support to this reading besides L&S's occasional explicit endorsement. The stronger claim seems to be the only one suited to their overall project and to make sense of their basic distinction: genuine meaning to the *convention* side; other stuff to the *imagination* side. Any chance of meaning outside convention would dull the sharp distinction that is promised.

We saw in § 8.1.2. that, for the most part, L&S simply reject for speakers' communicative intentions (in action) any role in determining the meaning of their utterances. This is their position with regard to ordinary occasions of verbal communication involving competent speakers and standard meanings. Occasionally, however, they withdraw to a slightly more accommodating attitude towards the idea that such intentions might contribute in some way after all. Here is a passage about the coining of a new term or the introduction of a new name:

> … direct intentionalism has to invoke special principles and special explanations to account for our judgments about creative language use. It's true in these cases that the speaker has intended to use her words to invoke something new in the world, and that this intention has been recognized by the interlocutors. But that in itself is not enough, on our view, to conclude that the speaker has succeeded in making a contribution to the conversation. This further conclusion depends on standards of meaning that describe not only the speaker's intentions but also interlocutors' broader background commitments to the communicative enterprise—commitments that govern the institution of new meanings as well as commitments to respect existing ones. (ibid., 224)

It is, however, still clear that, in their view, the intention to use words with a specific meaning, and its adequate recognition by the audience, are still "not enough" to make meaning. It all still depends on certain "standards of meaning" or "commitments" being in place. They ask us to consider, for example, "the commitments involved in securing the reference of a new proper name" (ibid.). They invoke Kripke (1972) to explain, very thinly, that "interlocutors must jointly commit to preserve the link between the name and its referent" (ibid.). When the matter is "a newly coined general term" (ibid.), to achieve "meaning depends on the joint commitments of interlocutors to triangulate a consistent, useful, and natural distinction between the things that satisfy the word and the things that don't" (ibid.).

They are not very elaborate about the commitments and social competences involved in these stages. They postpone more explanations to a later section but even there details are not abundant. I will leave the matter at this point for now but shall return to it afterwards. My purpose at this stage was simply to reinforce my reading of L&S's position, according to which they endorse i. and ii. stated above.

I reject both i. and ii. In this section I argue for the necessity of a notion of linguistic meaning that is compatible with the rejection of both i. and ii., one that would allow us to make good sense of the claims that linguistic meaning is possible outside conventions and standards, and that the adequate transaction of communicative intentions[170] can, by itself, generate meaning and linguistic communication.

This will translate into distinct positions and assessments with respect to a number of problems. It starts, of course, with the types of situations we have been paying special attention to: mistaken, non-standard uses of words that still do not cancel mutual understanding.

Contrary to L&S's position, I defend that the actually intended (in action) meaning behind faulty utterances, in those cases where that meaning is adequately recognized and handled by the audience, should not fail to be acknowledged a semantic status. We will have a genuine instance of linguistic meaning - of an especially important sort - whenever a speaker manages to satisfy her intention to get her words to be taken, by her audience, to mean what she wants them to be taken to mean, even in the absence of conventions and standards.

This is what happens with (199) according to the story. The psychological facts are not in dispute, and L&S agree that "it's clear to all concerned that the steward expects to deputize [her] hearer". The "speaker of (199) intends [in action] to contribute that the hearer is empowered to act in [her] place" (ibid., 218) - in particular, she intends her utterance of "jeopardize" to be taken to mean *deputize* - and this intention is not missed by the audience. As we saw, L&S do not think that this is enough for the interchange to be deemed of any semantic consequence. I oppose this assessment. I am convinced that there is a crucial sense in which "jeopardize", as it occurs in (199), means *deputize*, and also that, in general, given the right conditions, agents must be capable of generating linguistic meaning with a certain

[170] In this general category we can fit, in particular, Gricean m-intentions and Davidsonian semantic intentions.

amount of autonomy from prevailing standards and conventions.

The utility and application of the notion of intended, potentially non-conventional and non-standard, linguistic meaning that I am arguing for extend far beyond cases of faulty utterances. In the defense that follows we will encounter various such uses - accounting for word learning, describing the emergence and consolidation of conventions and standards, and explaining why agents participate in linguistic interchanges - where, in my assessment, we would want to say that there is linguistic meaning involved but a notion of standard and conventional meaning is not suited for the job.

As is perhaps clear by now, the relevant notion here - at least, the notion I have in mind - is none other than that of *first meaning*, as proposed by Davidson (see §7.2). To maintain that there is linguistic meaning whenever a speaker manages to satisfy her intention to get her words to be taken to mean what she wants them to be taken to mean - as I will be doing in these sections - is to maintain the pertinence and semantic quality of Davidson's notion. I will be recapitulating some of the points and reasons delivered in the previous chapter in support of its fundamental importance to language and verbal communication.

The dispute might perhaps be described as a terminological one over the extension of "semantic value" or "linguistic meaning", but it certainly amounts to more than a sterile quibble over words. Minimally, it reveals distinct theoretical options of consequence. In reserving the title for cases of standard and conventional meaning, L&S mark a sharp and fundamental division where, from my perspective, there is only a somewhat vague and marginally important one. At the same time, they refuse to acknowledge the fundamental continuity I am interested in between cases from both sides of this divide. They attach ultimate importance to what I find merely instrumental, however often indispensable, and they underestimate the importance of certain connections between language, mutual understanding and general interpretation that I find essential.

What I will be attempting in this section is not a complete and definitive defense of the *notion of first meaning*. Reflection and discussion on the idea of an intention-based notion of linguistic meaning, one that is, to some extent, prior or autonomous from conventions and community norms, has a long and rich history. Here I am not aiming at anything resembling an adequately representative account or survey of that history, nor of the problems and questions explored, and arguments and counter-arguments on offer. The consideration of all relevant contributions is clearly beyond the scope of these sections and of this dissertation. I

will confine myself to a brief and exploratory discussion of the topics and challenges raised by L&S against the viability and theoretical worth of a notion such as this, *first meaning*, as well as of the understanding of language and linguistic practices the notion entails, with intentions, mindreading and general interpretation playing a very heavy and central role in it. I will carry this discussion through four topics - the point of language, cooperation, foundation and innovation, and language acquisition - along the next four sections.

8.2.2.1. THE POINT OF OUR SOCIAL USE OF LANGUAGE

Throughout their book, L&S insist on an analogy between participation in the social practice of language and participation in games such as chess, baseball and golf (Lepore & Stone, 2015, 14, 222 and 254-5). They stress how the rules and scores in these games should remain firmly established and impervious to the participants' own intentions and understanding of those rules.

> … players who take up a game of baseball, golf, or chess, aren't just looking to agree at the end of the game about how the score evolved. They normally agree in advance on the rules that they will play by, the mechanisms and standards that they will use to resolve difficult or questionable cases, and even the sanctions that they will abide by when there is a breach in the rules. If we decide to play chess, what we've decided on in the first instance is that we are deferring to the rules of chess. We assume that we have a good idea what those rules are, but what we've agreed is to play chess rather than just to follow the rules that we know. So if one of us has an incorrect idea about the rules of chess, we will talk about it and look it up in a book. Whatever the rules of chess turn out to be is what we will play by - generally speaking, of course. (ibid., 254-5)

The same should be the case, that is, the intended lesson, for the rules and standards of meaning and conversation. Just like there are fixed rules governing one's moves when playing a game, there are fixed norms and standards that establish what is possible to mean and how. This is undoubtedly the case, to some degree. There is also, however, a decisive divergence between the two matters that I would like to point to in response to L&S's case.

The analogy is essentially flawed, I maintain, because the social uses of language, unlike moves in those other games, have always a point outside language, an external goal that motivates and explains such uses. It is the game's rules that, by themselves, determine the value of each move and position within the game and so, accordingly, justify and motivate the players' strategies and performances. It is not like that with language. Language

is ordinarily used with some ulterior end in mind that has nothing essentially to do with language, its rules, norms and standards.

The point of our social uses of language, articulated at its most general, is mutual verbal understanding among interlocutors. I team with Davidson when he asserts that "what matters, the point of language or speech or whatever you want to call it, is communication, getting across to someone else what you have in mind by means of words that they interpret (understand) as you want them to" (Davidson, 1994b, 120). All else, namely, conformity to norms and standards, weighs little in comparison, if not for its instrumental value in furthering that overarching purpose. As I see it, Davidson's analogy, and not that of games, is the one that illuminates the relation of linguistic norms to communication:

> Using a word in a nonstandard way out of ignorance may be *a faux pas* in the same way that using the wrong fork at a dinner party is, and it has as little to do with communication as using the wrong fork has to do with nourishing oneself, given that the word is understood and the fork works. (Davidson, 1994b, 117)

L&S insist that behind ordinary uses of the language there is more besides the communicative motives and intentions that, in each case, specify the general drive towards mutual understanding. I do not dispute or doubt that, quite commonly, there will also be, on the speaker's part, an effective desire or commitment to conform her use of the language to that of the community - that is, to follow the community's conventions and standards. We can even conceive various reasons behind this additional layer of intentions from personal motives, such as preferring to look educated and literate rather than ignorant, to some felt responsibility or obligation to a social norm, perhaps explainable by the need to contribute to ensure the long-term preservation of a broadly shared language and ease of communication even among strangers[171].

L&S press this point with respect to the steward's story and her utterance of (199). They concede that she "intends to contribute that the hearer is empowered to act in [her] place" but they stress that "[s]he also intends to contribute the conventional meaning of what [s]he said, which is in fact that the hearer is at risk" (ibid.). From this they even derive an impasse for the prospective intentionalist: "We seem to be equally justified in reporting the situation with either perspective" (ibid.). In such a case, which intention in action should we

[171] Interestingly, in this latter case, there is still subordination to the broader value of mutual understanding.

allow to determine the meaning of her utterance? The final step in their reasoning, unsurprisingly, is to take such potential impasses to recommend *direct* instead of *prospective intentionalism*. To escape “inadvertently conflicting commitments” (ibid., 218), they claim, “we must appeal to our standards of meaning to pin down the speaker’s contributions” (ibid.).

As I have said, I do not dispute the existence of such intentions and commitments What I dispute is their ability to rival in vividness and relevance the speaker’s communicative motives and intentions, for the privilege of determining the content of her utterance. If several commitments and intentions can determine our actions, we must also admit that they will hold different weights, different degrees of importance and priority. My claim is that central and foremost, ruling supreme over the social practice of language, figures the goal of understanding and communication. That, I believe, is the main line of business in the social use of language; all else is secondary.

This would allow us an easy escape from alleged impasses of the type presented above: simply privilege the main intention, the communicative one, and sacrifice the background commitment to social norms and conventions. Take our central example once again, (199). Plausibly, only one of the competing intentions would have decisively moved the speaker to produce her utterance. I mean, of course, the intention “to contribute that the hearer is empowered to act in [her] place” (ibid., 218). Very rarely, if ever, will people speak ultimately, or principally, for the sake of conforming to the community’s conventions and standards.

The view I am proposing has intuitive force and credibility. Further reflection and discussion - perhaps even empirical research - would be needed to further illuminate and, eventually, decide the matter. The ambition here is simply to sketch what I think is a promissory reply to the obstacles raised by L&S, not to settle a definitive answer. In the face of their adverse arguments and imagery, I remain convinced that an intention-based notion of linguistic meaning such as *first meaning* is both viable and needed. The threatening prospect of having the social use of language drift into a ruleless and hence pointless practice, or of having interpretation hindered by the sort of indecisions that L&S invoke above, seems less likely once certain distinctions become patent. Unlike a ludic, self-contained, activity, with its internal and fabricated values, the practice of the language responds to values outside itself. In this specific context, some of those values - namely, mutual understanding - are

flagrantly more important and mobilizing than others - namely, commitments to social norms.

8.2.2.2. COOPERATION

L&S defend that the view they oppose, Prospective Intentionalism, is one that is excessively dependent upon the effectiveness of cooperation among interlocutors. They hold that this speaks against the view since they are convinced that no form of conversational collaboration is generally necessary for an utterance to have a meaning.

They consider a couple of forms of collaboration - "practical collaboration" (ibid., 220) and "agreement on the issues under discussion" (ibid., 221) - before they reach the final candidate, a shared commitment to "reaching mutual understanding" (ibid.). As far as I can tell, this is the only real contender when the question is how or whether to characterize meaningful public uses of language as episodes of collaborative interchange among interlocutors.

In response, I must start by noting that the argument and its intended conclusion are not in actual contradiction with the modest position I am defending here. I believe that the regular commerce of meaning intentions (in action) requires collaboration among interlocutors, but the point that I am defending in these sections is simply that linguistic meaning is often determined by speakers' intentions in action, not that all linguistic meaning is immediately dependent upon such intentions. As noted above, I am ready to acknowledge other interesting notions of linguistic meaning. Since I am arguing for the possibility of meaning in accordance with this model, and not against the possibility of any meaning outside it, the question of alleged exceptions presents no actual threat.

Furthermore, even a more ambitious claim could arguably hold its ground against the proposed exceptions to universal collaboration. The examples L&S put forward are hardly persuasive that genuine conversation can in fact ever be accomplished without some minimal communicative collaboration. I can accept that there will be some borderline cases where, although interlocutors share no interest in understanding each other and make no genuine attempt at communication, their utterances will still, in some sense, carry some form of

linguistic meaning. Take L&S's example (ibid., 221) of Alice's utterance, a deliberately incomprehensible piece of instruction[172]. However, I find cases such as this far from sufficiently exemplary to allow or support any conclusion regarding what is and what is not "essential to conversation" (ibid., 222). Interchanges of the sampled type strike me as deviant specimens at most, more like perverse imitations of genuine conversations than the real thing.

In sum, I do not think that these considerations and doubts regarding collaboration can significantly weigh against the claim I am here defending. Successful interchange of intended meanings - of which first meaning is a form - requires minimal communicative collaboration, and the fact that some remote uses of language do not rely on such collaboration, as far as I can see, cannot be turned into a serious threat to the semantic quality of such interchanges.

8.2.2.3. FOUNDATION AND INNOVATION

In my view, we also need the notion of *first meaning* to account both for *language fixation* - that is, to explain how it comes about, or what makes it the case that the expressions in a language mean what they mean - and *innovation in the language* - that is, to explain how it happens that new expressions are introduced in the language, or that old ones become associated with different or new meanings.

There is an obvious but important sense in which first meanings must precede and pave the way for any form of stable expression meaning. Any plausible account of the emergence of standard and conventional meanings appears obliged to appeal to prior antecedent cases of successful interchange of non-standard and non-conventional first meanings. Interlocutors must first, somehow, manage to coordinate on the meaning of some

---

[172] "As a final case, we turn to problematic cases where interlocutors are not even committed to reaching mutual understanding. Such interactions are certainly attested - much of the humor of Scott Adams's workplace comic strip Dilbert turns on the frustrations of such cases. In one strip, partially transcribed in (200a), Alice, a representative from technical support, tells Tina the procedure to resolve a computer problem. Tina sheepishly admits that she can't make sense of the instructions, but Alice acts as though she has done her best to explain the fix.

(200) a. Alice: Just disable the local cache mode to fix the MAPI settings and delete the duplicate messaging subsystem registry key.
b. Tina: What if I don't understand anything you said right then?
c. Alice: Good grief! I can't make it any simpler.

Subsequent developments confirm our suspicion that Alice's efforts in the conversation are at best halfhearted. We see her walking away, thinking to herself:
"It's funny because it's cruel."

expression before that particular type of use, the particular association of meaning and expression, can become standard or conventional. Conventions arise as the rational propagation of such, at first, non-conventional uses. Standards are constituted by abstraction from a large enough number of such, at first, non-standard uses.

I do not find an alternative proposal in Direct Intentionalism about how languages are fixed. L&S do not appear particularly impressed by the problem - certainly, it is not moving them to reconsider the role of speakers' minds and intentions in semantics. They seem committed to the view that these fundamental steps occur outside semantics proper. I have strong reservations about such an option. What decisively distinctive new features can such uses acquire from repetition and proliferation? What could justify that the derived uses be granted the semantic status that the original ones are denied? I do not think there is a good positive answer to these questions. In particular, I firmly resist the idea that the question could be resolved in favor of Direct Intentionalism on grounds of better coordination or understanding. As abundantly illustrated by now, it often happens that coordination and understanding are reached outside standard and conventional uses, just as standards and conventions are no guarantee against incoordination and misunderstanding.[173]

Furthermore, there is also linguistic innovation. We must not neglect the fact that natural languages, even as commonsensically individuated, are in permanent transformation. What could the vehicle of such changes and innovations be if not instances of the successful interchange of non-conventional and non-standard intended meanings? Here, then, is the second function, intimately related with the first, that, I defend, requires a notion such as *first meaning* that is able to easily bridge between the psychological and the semantic.

With regard to this second theme, L&S have much more to say. Naturally, they agree that languages change and that new and revised meanings emerge. They also hold that their proposal is well equipped to deal with and explain semantic innovation. In fact, they hold that it fares better than prospective intentionalism in that at the same time as it acknowledges creation and innovation, it also allows us to better distinguish creative from conventional uses.

[173] The utterance of (199) is a case in point. Coordination and understanding are not in danger, since the intended meaning is "clear to all concerned" (ibid., 218). Unless, that is, one were to take the uttered words at their standard value. It is the latter option that, if just by itself - that is, unaccompanied by further investigation, correction, retraction and so on - would break communication and hinder successful interaction.

For my part, I fear that they impose conditions that are too strict on what they are willing to take as legitimate additions and innovations. On the one hand, we find various passages affirming the need to acknowledge and account for the invention and improvisation of meanings. On the other hand, however, their insistence on conventions and other pre-established solutions for coordination appears to constrain invention and improvisation to take place only in accordance with specific rules and established procedures of invention and improvisation. It is full-fledged instances of *improvised improvisation* and *invented invention* that their proposal seems to have trouble accommodating. In the rest of this section, I will try, first, to make their position on the possibilities of semantic innovation clear and, second, to show where I think it falters.

The restraint in L&S's concessions to meaning improvisation is made explicit in a number of passages. It is present, for instance, when they claim that "if much of meaning is in fact improvised, then we will always need to describe institutions that let us make meaning, not just institutions that let us use established meanings" (ibid., 242-3). There is also the passage, already quoted above, affirming the need for "social competence" and "mechanisms of a specific type" behind "improvised meaning":

> Our theory of semantics has to be compatible with universals of meaning, with *improvised meanings*, and with our practices for disambiguating, refining, and negotiating meanings. ... These aspects of meaning are still manifestations of interlocutors' social competence, we will suggest - but *it involves knowledge and mechanisms of a specific kind.* In particular, interlocutors are normally coordinating on a process of inquiry, through which they commit to make their meanings public. *Semantics*, we will suggest, *describes the social competence that specifically supports this coordinated inquiry*. (ibid., 245; emphases are mine)

Indeed, L&S seem to be reaching only for a qualified idea of improvisation. Even with regard to meaning improvisation, they stress that we "[make] meaning in established ways".

> In conversation, normally, each of us defers to a broad set of background commitments about meaning. These commitments describe how to update the record, but they also describe our responsibilities in using language in accord with the broader community, in introducing and negotiating new meanings in special situations, and in working to make sure that we understand one another as *making meaning in established ways*. Interlocutors have an interest in understanding one another, but also have broader interests in using their understanding to pursue consistent and public meanings. So while we want to coordinate on the record, *we do that by agreeing on a wide range of additional*

background commitments that cover the full range of conversational dynamics. (ibid., 255)

Besides their statements directly on the topic, there is the question of how to fit improvisation and invention with some other views and positions of theirs. Again, these other views and positions appear compatible only, at most, with qualified and watered-down conceptions of invention and improvisation. I have in mind much of what we have been considering, for instance, their rejecting the need for mindreading in meaning-making, their strict diet of basic intentions in semantics, and their understanding of cooperation among interlocutors as inessential to communication. All these elements converge, as we have seen, on L&S's stance on *faulty* utterances such as (199) that I expounded above. Tellingly, they never revise it. In the midst of their treatment of improvised meanings and the conversational record, they return to that malapropism only to reiterate that the utterance should be ascribed its standard meaning. They decline the opportunity to present the malapropistic utterance as an acceptable case of improvised meaning when the example re-emerges to illustrate the problems with a private understanding of the record.

> Take malapropisms. Interlocutors often expect not only to be getting their meanings across, but to have expressed their meanings in ways that are consistent with the usage of a broader community. These expectations are sometimes wrong, as we saw with (199) … What happens in such cases can vary, of course. But one thing we can do in response is to prioritize our commitments to use meanings in line with the broader community in tracking and shaping the course of the conversation. (ibid., 253)

Of course, the record would be useless if one were to identify the record with one's representation of the record. It is fair enough to claim that the "record, in this sense, is not exhausted by our ideas or intentions for what we contribute with our individual utterances" (ibid., 253). But indulgence to solipsism is not the only way to allow for a non-standard ascription in examples like (199). What is interesting in these cases is that we are assuming the hearer to be able to recognize the intention (in action) behind the speaker's words. Hence, there is actual understanding; there is actual coordination or agreement on the intended meaning. So why would we refuse, as L&S argue we should, that in these types of cases interlocutors attain actual coordination on the conversational record?

Here they make explicit some of the ways in which, according to their proposal, interchanges fall short of inscription onto the conversational record and, consequently, of semantic status:

> We contrast updates to the record with other cases, where an invited inference is suggested, indicated, or *revealed by broader background knowledge about agents and their beliefs, desires, and choices*. These latter inferences are not grounded in the kind of social competence we have considered, because it is not agents' mutual expectations about one another that are decisive. (ibid., 256, the emphasis is mine)

The intended meaning of (199), *deputize*, is indeed "revealed by broader background knowledge about agents and their beliefs, desires, and choices". As such, L&S claim that it must be excluded from the conversational record. This is their verdict - or theoretical option - but I still find no compelling reason behind it.

Perhaps the question is that L&S are convinced that in cases such as (199) interlocutors can never be sure. Perhaps they fear that, without strong social institutions backing it up, coordination will always be precarious and uncertain. If that is so, they clearly exaggerate the power of such institutions and the congruence between the two distinctions. When we move from the general principles to the actual cases, we are forced to recognize that the divisions are not so tidily coincident. Frequently, coordination is much more solid outside the mechanisms and institutions specially dedicated to ensure the publicness of meanings. For instance, agreement and coordination seem much more solid over the intended meaning of the steward in (199) than over the conventional meaning of Alice's words in (200).

There is also the suggestion that a prospective intentionalist would be committed to a painstaking, rich and complex process of interpretation for each utterance.

> Grice (1957) goes so far as to say that conventional meaning itself is just a generalization about what people tend to intend when they use words. In other words, we should take the eclectic interpretive reasoning we need for creative language use as the model for how communication normally proceeds. (ibid.223)

This, however, does not follow from admitting more than basic intentions into semantics. Stephen Neale (1992, 551-2), for one, emphatically alerts us against the distortion present in this understanding of Grice. There are many ways to recognize a communicative intention. Ordinarily, interpretation will simply run fluently with words being taken at their standard value, with no call for elaborate inquiries. A tacit judgment of no detectable incongruence, of no sensible signs of misunderstanding, will ordinarily be all that is expended, by the interpreter, in confirming a meaning ascription.

L&S elaborate only one example of meaning improvisation that actually gets onto the conversational record - the baptism of a dog that is to be called "Luna" (ibid., 223, 257-8). What they stress is that, despite there being "no specific convention in advance governing the use of the name" (ibid., 257), and our consequent inability to "describe interlocutors' coordination on the record in terms of their antecedent mutual expectations" (ibid., 258), there are still other social mechanisms and background commitments or mutual expectations governing the event. That, according to them, is what makes it the case that "the speaker and her audience are still coordinating in this case" (ibid.258). They borrow from Kripke's (1972) original insights to briefly characterize the norms of conduct for both speaker and audience.

> The speaker must have a specific pairing of form and meaning in mind with her use of the word Luna. She must be able to specify and clarify this meaning enough for the purposes of the conversation, and she must be acting according to a consistent pattern that can serve as a precedent for referring to Luna as Luna across a broader community, to the extent that's practical. The audience, meanwhile, in borrowing the speaker's reference to her dog as Luna, must commit to understand and pursue the meaning of the word systematically throughout the conversation. (ibid., 258)

I have two things to say here. First, even this single example of the most propitious sort seems to fail to support direct intentionalism. Second, the example is not representative of the diversity of situations and methods through which new and modified meanings emerge. I will consider each point in turn.

I believe the Luna example is compromised in two ways. To begin with, I think we must question the *convention-like* character of the described *background mutual expectations* governing the event - those mutual expectations that should earn the episode its place on the conventional side of L&S's divide. L&S suggest that there is some decisive affinity:

> These particular mutual expectations seem quite general of course. So we needn't say they are conventions, in Lewis's sense, to the extent that they represent general constraints of the human language faculty. However, as always with the dynamics of the scoreboard, we can certainly imagine alternatives to them. Together, then, these mutual expectations show how interlocutors are committed to collaborating on the record in this case. (ibid.)

They are not very detailed either in the characterization of those mutual expectations or in the elaboration of what exactly constitutes their distinctive, *convention-like* character that is meant to allow the new name to figure on the record. They mention the possibility of *alternatives*, which is an important feature of conventions. This would hint at the presence of

some arbitrariness in the deed, and the consequent need for specific and learned knowledge or competence. I am convinced, however, that there is actually little room for alternatives - genuine, equally suited alternatives - to the strategies described. How could it happen in any other way? Could the speaker instead not "have a specific pairing of form and meaning in mind", fail "to specify and clarify this meaning enough for the purposes of the conversation" or to act "according to a consistent pattern that can serve as a precedent for referring to Luna as Luna across a broader community"? Could the audience do anything else but "commit to understand and pursue the meaning of the word systematically throughout the conversation"? If not, what alternatives are they referring to? I am convinced that the protocol in such cases is not really a protocol, but rather the only rational course available for speaker and audience, intrinsically preferable and thus dispensing learned social competences and prefabricated solutions. The challenge is extensive to all the "wide range of additional [that is, not-exactly-conventional] background commitments" (ibid., 255) in what concerns this question of meaning improvisation. I believe that a stronger case is needed to justify their special status and the central division in the book that depends on that assumption.

Besides the general mutual expectations repeated in all situations of the same type, more is needed for the introduction of the new meaning to be successful. It is the speaker's intentions in action that specify the new name's intended reference, and no familiarity with the codes and norms of language could dispense the audience from having to employ their mindreading abilities in figuring it out. General mutual expectations about how the introduction of a new name works can only take us so far. Interlocutors will also be resorting to their "broader background knowledge about agents and their beliefs, desires, and choices" (ibid., 256) when they manage to reach verbal understanding and fix a name just by exchanging "...and this is Luna". Only that can allow them to figure out that the speaker is referring to the dog, not to a temporal stage of it, not to an undetached part of it, not to its breed, and so on; that it is a new dog, not an old one that has been rebaptized, or one that the hearer is being reintroduced to after years of being apart after which it has changed enormously, and more.

The second complaint about L&S's sole example is precisely its soleness. The example is not representative of the diversity of situations and methods through which new and modified meanings emerge. Robyn Carston issues a similar criticism. In particular, she notes, "[u]naccountably, L&S omit all mention [to a second] source of new linguistic

meaning conventions" (Carston, 2016, 621). She is referring to "pragmatically based-innovations", where the sort of features that L&S emphasize as crucial for coordination and defend to be still present in the *Luna* example - pre-established mutual expectations, protocols, arbitrariness, specificity of the relevant knowledge and competence - recede even more dramatically, or disappear completely. Carston offers three classical examples, both plausible and banal, of pragmatically-based lexical innovation:

a. The boy porched the newspaper.

b. She managed to wrist the ball neatly over the net.

c. The prisoner houdinied his way out of the locked cell. (ibid., 620)

What is important here is that an adequate response to these utterances requires the mobilization of resources outside what L&S strive to circumscribe as specifically linguistic knowledge and competences. The crucial role of communicative intentions, mindreading, background knowledge and general intelligence in interpretation is much more evident in these types of cases.

Finally, there is the case of unintended malapropisms that are interpreted as intended by the speaker. I do not see why they would be refused a place in the same continuum we have been exploring. We must not, and need not, mask the differences among the various cases of innovation - more or less protocolar and explicit, more or less by the book, voluntary or not - but the crucial similarity is also beyond doubt: in all these cases a new use of a term is shared among interlocutors; everyone agrees on what it is expected to mean in the utterance in question and that is all that is needed for verbal communication to be effective.

I find it reinforcing to notice that even such unintentional misuses can gain currency and stabilize. It is a contingency if the community does not pick up (199) as a precedent for a new convention; it would not be the first linguistic convention built on an error. Gareth Evans, in response to Kripke, noted that the name '*Madagascar'* had initially referred to a certain portion of the African continent, not to the island.[174] The change in reference took place unintentionally, and the current conventional reference of the name is founded on a mistake. There is also the famous example of *'livid'*. The term, in time, came to mean also

[174] See Kripke, 1972, p.163.

reddish and *flushed*, while originally it only meant the opposite - *bearing the colors of a dead body, blue, grayish-blue, whitish.*

L&S's stance on semantic innovation and improvised meanings is consistent with the rest of their proposal, and so is my appreciation of it. Once again, I do not think they manage to adequately justify - and I actually disagree with - their choice of the limits and categories with which to describe and understand language, meaning and linguistic practices. As I see it, the crucial distinction is not whether or not there are institutions and regularities in place supporting coordination, the essential question is simply whether or not interlocutors manage to attain coordination. The regular dialogue within the bounds of convention, the coining of a new name or term, the successful use of an ingenious and motivated neologism requiring informed and intelligent deciphering, the infant's learning of her first words[175], the unintentional malapropism that still permits the interpreter to get what the speaker actually intended (in action) to convey, all these cases are similar with respect to that fundamental aspect: the speaker's "intention to be taken to mean what [she] wants to be taken to mean" (Davidson, 1994b, 120) is satisfied.

8.2.2.4. LANGUAGE ACQUISITION

There is also the closely related issue of language acquisition and, in particular, word learning. Just like there seems to be no way to fix a conventional or standard use that is not built over successful cases of interchange of elementary first meanings, there seems to be no way of *radically*[176] learning the language of our interlocutor if not by aiming at the first meanings of her words. In previous chapters I followed at length the child's and the radical interpreter's path to the meaning of words in the relevant language and defended that it must necessarily pass through the discovery of the speaker's communicative and semantic intentions. The speaker's use of her words is the sole clue available to their meaning. Child and interpreter must try to infer the speaker's intentions and other thoughts behind such uses so as to get to their first meaning ascriptions. The meanings thus targeted are the meanings intended by the speaker and it makes no difference at this stage whether her use of the words

[175] Once again I find it opportune to note that all these cases exhibit a flagrant affinity to those original innocent situations where our first words were learned, and where, for sure, the interpretative enterprise had to proceed unaided by any type of previous agreement or social institution - at least for a while.
[176] That is, without the benefit of a second common language.

corresponds to a widespread practice in the community. In the same spirit as above, I object to stipulating a fundamental breach between what speakers start by aiming to learn - intended meanings - and what, eventually, they ultimately reach - intended meanings that are also standard.

Besides, even mature speakers rest far from immaculate competence regarding a vast number of expressions of the language. Externalist insights made this fact all the more vivid. The learning is continuous through life and it will frequently be a matter of figuring out the meanings that one's interlocutor intends for her words, the interlocutor's first meanings, ordinarily trusting that they will correspond to the community standard.

For an illustration, let us return once again to our patient in the arthritis story. Imagine that he has just started to describe his symptoms to the doctor. He begins by saying he has a pain so-and-so in the thigh area, that it has been going on for weeks now, that it seems to get worse in such and such circumstances. Then, just before he opens his mouth to tell the doctor that he believes that what is affecting his thigh is, once again, arthritis - both T and the doctor are perfectly aware that T is very prone to arthritis and in fact has arthritis in several other places - his doctor speaks. He says:

- Good thing you have no joints there. Now you see, there are other ailments besides arthritis.

This utterance, besides confirming his doctor's awkward sense of humor, gives him the opportunity to improve his knowledge of arthritis and, crucially to our case, his understanding of the term. He should easily infer from the doctor's words that arthritis can only happen in the joints and, hence, that he cannot have arthritis in his thigh. Naturally, this will prevent him from formulating the famous utterance expressing a fear or belief that he no longer holds.

T's discovery clearly involved mindreading. It required some mild delving into the doctor's intentions, expectations and beliefs underlying the doctor's utterance. The doctor had assumed T was aware of the fact that arthritis can only occur in the joints. That is why T would find it good not to have joints there, with arthritis being something to be dreaded and his being so prone to it. Some other condition is to blame for T's discomfort.

Of course, there is nothing special or original about this example. It just serves to illustrate how an externalist understanding of speakers' relation to language is at least just as

propitious for asserting the indispensability of mindreading and rationalization in interpretation: it is (also) because externalists are right, and linguistic agents, in some sense, know it, that interpreters exercise a permanent and (for most of the time) almost effortless alertness in conversation for the kind of event that would justify any addition or revision to their knowledge of the language. Together with the imperfection of speakers' linguistic knowledge, externalism stresses the triviality of episodes prompting speakers to revise their knowledge of the language.

Situations of this type are fairly common and this shows that the need for a notion of *first meaning* is not confined to a very initial moment of the speaker's contact with language. Speakers' attention to their interlocutors' semantic intentions span the whole of their linguistic career, and it is not only determinant to communication but it contributes, at each moment, to their conception of the standard or community meanings.

In their reply to a group of commentaries, and addressing some of Robyn Carston's points in particular, L&S admit that "there's powerful evidence that intention recognition is crucial for language learning ..., for understanding creative language use ..., and in recognizing the information that speakers do not encode, but simply reveal" (Lepore & Stone, 2016b, 652). However, they hold that "[t]his is compatible with [their] view", because it "does not establish ... that intention recognition also figures front and center in semantics" (ibid.).

If I understand them correctly, they do not dispute the role of intentions in all these achievements, what they reject is that these topics, and the elements and processes involved therein, are proper objects of semantics. In particular, they concede that we should not fail to consider speakers' intentions and learners' attention to them when studying how people learn languages, but not when the question is "What is the meaning of some competent speaker's utterance?". I accept that the questions and concerns are different in each case, and that it is in general wise and recommendable to preserve some boundaries and to distinguish disciplines. However, I still insist on the necessity of recognizing an intimate relation between the two subjects and inquiries here at stake. I believe it vital for semantics to operate with notions that are suited to be applied in the understanding of what takes place when people learn their languages. The danger here is that in refusing an intention-based notion of linguistic meaning, one that incorporates both conventional and non-conventional, standard and non-standard, meanings - a notion like *first meaning* - L&S seem to be relinquishing the

type of notion that could actually fit what people aim at when they learn a language. Different disciplines should keep their distinct identities, but cooperation and natural continuities among them must not be compromised.

8.2.2.5. CONCLUSION ON INTENDED MEANINGS

In these last sections, I have put forward my case in favor of a notion of intended, potentially non-conventional and non-standard, linguistic meaning, such as *first meaning*. In dialogue with some of the positions and arguments raised by L&S, I have presented the reasons behind my conviction that this notion is both viable and valuable in the study of meaning and language.

I have argued that the obstacles, identified by L&S, to a conception of linguistic practices and interpretation based on this notion are not insurmountable.

I have defended that we can easily accept and accommodate speakers' standing intentions to follow the norms and standards of the community alongside speakers' specific communicative intentions. Their coexistence will hardly obstruct the determination of meaning even in cases of disagreement between these intentions. Some intentions are more important than others, at least relative to some particular understanding or description of the action and event in question. Meaning something, I proposed, should not drift too far from speakers' communicative intentions.

I also responded to L&S's fear that a view of linguistic communication based on an intention-based notion of linguistic meaning would be unable to adequately distinguish conventional from creative uses of language. I defended that the implication of a natural continuity between both forms of meaning - when coordination is in any case reached - is a merit, not a fault. Furthermore, a shared intentional genus still allows us to distinguish various species whenever pertinent, namely, intended and conventional versus intended and non-conventional meaning. I also explained why I do not think that acknowledging intentions a more substantial role in semantics would, by itself, compromise coordination or represent interpreters as committed to unnecessary toil.

I insisted that cooperation, at the basic level that is relevant to the discussion, is indeed very pervasive. I expressed some reserves about the soundness and representativeness of L&S's examples of uncooperative conversations.

I also noted that the recognition of an intention-based notion of linguistic meaning does not preclude the possibility of other interesting notions of linguistic meaning. Surely we do not have to give up the idea that, in some sense, the steward used an English term, "jeopardize" that is standardly taken by English speakers to mean *jeopardize.* My principal contention in these sections is that this is not the only notion of linguistic meaning that we can and should keep.

Besides viable, I also tried to show that the notion of first meaning is extremely valuable. Without a notion of intended, potentially non-conventional and non-standard, linguistic meaning, we would be deprived of a semantic notion with which to address a number of questions: What, first and foremost, motivates our social uses of language? What makes linguistic interchanges a tool for human understanding? What brings about and constitutes the standards and conventions of meaning? What supports semantic innovation? How do people learn their languages, especially the words? Standard meanings will occasionally - and, to some extent, accidentally - fit the job, but frequently they will not - namely, when standards and intentions fail to coincide. I have presented my case defending the claim that *first meaning* is the notion we really need in describing and understanding all these matters.

8.3. CONCLUSION

My principal goal, in these last two chapters, was to defend that the prevalence of conventions, stable standards, shared norms and uses, does not change the fact that linguistic interpretation requires rationalization. In this chapter, I addressed a very recent proposal, L&S's direct intentionalism, which articulates a particularly radical form of opposition to that idea.

In §8.1, I started with a general characterization of L&S's direct intentionalism. Some original elements and ideas had to be explained in some detail. The divergences between direct intentionalism and the conception of meaning and communication endorsed in this dissertation become especially evident with respect to a specific class of utterances - faulty utterances - which were given central stage at §8.2.

In §8.2.1. and subsections, in opposition to direct intentionalism, I tried to show that linguistic interpretation in accordance with the prevailing norms and standards is inadequate

for faulty utterances. On this basis, I generalized and argued that the social environment is never enough to determine the meaning of some utterance, and that the speaker must also contribute by doing her part. The adequacy of even a standard interpretation is conditional on the speaker's fulfillment of some conditions: she must be minimally competent with the expressions used; she must also hold compatible states of mind, that is, she must want to mean, expect to mean, and intend to mean that standard meaning by means of the utterance in question.

I have accepted that, in some sense, a word might be said to preserve its meaning in whatever circumstance, independently of the mind and context behind its utterance. However, I have stressed, that would correspond to just one particular notion of linguistic meaning, one with quite limited application and elucidative potential within the ambit of regular contexts of linguistic practices. What is more, this form of preservation of meaning has nothing to do with participation in conventions or in the social institutions of externalism. As we saw, §8.2.1.1, an agent cannot count as a member of the population in which some convention is in force if she is ignorant of that convention. Nor - as I argued for in §8.2.1.2 - will a speaker be able to count as putting forward a certain externally individuated content just by uttering the communal term if she is thoroughly incompetent with it. As noted above, "jeopardize", from our steward example, will not be preserving its *jeopardize* meaning in the same way in which "arthritis" preserves its meaning in Burge's example. The purpose cannot be understanding and communication if one insists on committing the speaker to a meaning that, however standardly associated with the expressions she has used, we know she did not want, expect or intend to mean. A standard interpretation of a faulty utterance cancels communication and understanding.

Afterwards, especially from §8.2.2. onwards, a number of other types of situation were identified where any operative link between the words and a standard meaning is missing. In some cases, there is a standard use or meaning for the expression in question but that regularity is for some reason interrupted or irrelevant for that particular linguistic interchange. In other cases, there simply is no standard at all. Besides the original examples of intentional and unintentional malapropisms, we had the chance to note a fuller range of forms of semantic deviation and innovation. The dynamic character of language is routinely attested with new names and new general terms, along with new uses for old terms. These may be in force just for the duration of the conversation, they may stabilize into a broadly

shared community convention, or they may end up any place else in between.

Then there is also the context of language acquisition. Logically, to some degree, learners will ignore the standards or conventions associated with the expression being learned. We have already explored this idea in previous chapters where we focused on an agent's initial steps into the knowledge of some language. In §8.2.2.4. I expanded these considerations by stressing the fact that speakers are permanently learning their languages. Our incursion into the themes and questions of semantic externalism granted us the opportunity to notice, with increased accuracy, the scale of the ordinary limits of our knowledge of meanings. Commonly enough, it is the interpreter himself who lacks some relevant piece of knowledge concerning the standard meaning of some uttered expression. Every occasion potentially holds the chance for advancing that knowledge.

What follows is the general insufficiency of standards and conventions for linguistic interpretation. Since faulty utterances and other forms of non-standard uses of the language are a real, and often actualized, possibility, linguistic interpretation cannot take the form of an uncritical and almost automatic operation of decodification of expressions in accordance with a stable set of uses. When communication and understanding are the question, the interpreter cannot allow himself to be exclusively guided by the conventions and standards of the language, inattentive and indifferent to the speaker's mind and competence with the terms employed.

Even prior to any question of reinterpretation, the mere decision over the adequacy or inadequacy of a standard or expected interpretation, when there is one, requires an investigation that already transcends the simple application of the linguistic knowledge in stock. Notice that - as explained in the previous chapter, §7.3.2. - it is not only when things go badly that the need to go beyond conventions and standing linguistic knowledge arises. The sheer possibility of detecting that something is wrong can only be secured if some form of prior vigilance is exercised. It is that prior vigilance that will allow the interpreter to issue a necessary judgment - even if tacit, most of the time - either in favor or against the adequacy of the standard interpretation. We need it both to detect that something is wrong with a *jeopardize* ascription in (199), but also to confirm the normal and expected meaning when things run smoothly.

That vigilance takes the form of rationalization. Interpreters must be permanently assessing the coherence and intelligibility of speaker's states and actions. They must be

continuously testing the rational match between, on the one hand, what the interpreter knows or thinks he knows about the speaker, her general condition and mental life, and, on the other hand, the meaning he is considering ascribing to her words.

There is room for distinctions here. Obviously, the kind of testing that *silently* takes place in the normal and smooth course of *interpretation as expected* will be less demanding, and harder to detect or recover introspectively, than the kind of investigation that is prompted once some deviation is detected at that basic level. Nonetheless, it is still testing, mindreading, and rationalization - I do not think we can make sense of the occasionally detected need to interrupt standard interpretation without accepting this.

All that has been mentioned in the previous paragraphs of this conclusion allows us to reiterate the first of the two claims distilled in the previous chapter regarding the role of conventions in linguistic communication. The permanent possibility of non-conventional and non-standard uses entails the general insufficiency of speakers' knowledge of meaning conventions for linguistic interpretation. This already secures for me what I need from this chapter for the general thesis of this dissertation - that interpretation requires rationalization. The mere interpretive decision over whether or not to go with the standard interpretation, when there is one, already commits the interpreter with the continuous effort of rationalization.

However, if I am right, the role of rationalization in linguistic interpretation does not end here. On top of the first interpretative layer just considered, there is also the rationalization and mindreading involved in the actual process of reinterpretation or creative interpretation when a standard interpretation is not available, or is concluded to be improper for the occasion by the interpreter. If the interpreter is to arrive at the actually intended meaning, instead of just stopping at the cancelation of the standard one, this will require the further exercise of his mindreading and rationalizing abilities.

The discovery of an intended meaning is, by definition, an instance of mindreading. Furthermore, in such moments where no standard or convention are available to help, the interpreter will be faced with an effective question demanding an intelligent solution. Ordinarily, in such cases - except, perhaps, on those situations where the speaker finds a direct way to declare what meaning she has in mind - the rationalizing effort will be intense enough for the discovery to be actually experienced as such. Freedom and ingenuity are essential here. The interpreter will have to work out his discovery of the intended meaning,

sensibly engaging with an indeterminately long collection of data, selecting elements of possible relevance and finding them a place in the right inferences, guided ultimately by the presumption of a shared rationality between himself and the speaker.

In §8.2.2, I recapitulated and added to the number of situations where some form of deeper interpretation unsupported by knowledge of meaning conventions or standards is necessary to reach the intended meaning of the uttered expressions. I argued that the notion of meaning implicated in these various cases is a fundamental notion of linguistic meaning, and that it corresponds to the Davidsonian notion of *first meaning*. I defended that no general proposal in semantics should neglect its role and importance.

That sometimes the intended meaning of some uttered expression can be discovered by an interpreter without the benefit of previous knowledge of the relevant convention or standard is, I believe, uncontroversially true. The question that actually lingers between me and the direct intentionalist is whether this should be taken to amount to linguistic interpretation and linguistic communication proper - in other words, whether a merely intended, non-conventional, non-standard meaning of some expression that is adequately recognized and handled by the audience should be acknowledged a genuinely semantic status. This is a theoretical question, I noted above, concerning what we take or expect our semantic theory to be. L&S give it a negative answer. I defended my positive answer by pointing to the indispensability of such a notion of intended meaning to the understanding of many core aspects of our practice of the language. A positive answer here entails the second of the two claims of the previous chapter: occasionally, meaning conventions are also unnecessary for linguistic communication.

General Conclusion

In this dissertation I have unfolded and defended a claim, that interpretation requires rationalization, and the general conception of linguistic meaning and communication that supports it.

In Part I, I have introduced and explored the notion of rationalization in the conceptual environment in which its need and presence is most evident, that of the interpreter's total and flagrant ignorance of the language of his interlocutor. Radical Interpretation is the thought experiment Davidson uses to reflect on these cases. In Chapter 1, I have explained it in detail, with the emphasis there being placed on Davidson's intense reliance on the Principle of Charity. The principle commands the interpreter to assume that the speaker is a rational agent. A great deal is contained in this instruction, especially because the notion of rationality here at stake is a very broad and a very rich one. The radical interpreter could not go far with only a minimal notion - often the one at work, explicitly or implicitly, in discussions on this topic - of conformity to the norms of formal theories such as logic, decision theory and probability theory. We need a much more inclusive, flexible and vague conception of rationality. As hard to grasp, contain and define as it may be, still the notion is neither vacuous nor idle. In particular, it plays the central and indispensable role in allowing me to do and affirm the things I do and affirm in this dissertation.

We need this notion, first of all, to explain how it is that the linguistic interpretation of a speaker must be integrated in the larger task of general interpretation of the whole person - an agent and a thinker, as much as a speaker. There is no chance of figuring out just the meaning of the speaker's utterances while not embarking on the consideration of the remaining aspects of her rational condition, as meanings, thoughts and actions mutually support and reveal each other. This is a theme to which I return time and again throughout the whole of this dissertation.

The mere identification of an interlocutor as a rational being already comes with the assumption of some very general and structural traits of her mind and conduct. We saw how the radical interpreter is expected to make use of these fundamental expectations about how actions, thoughts and meanings cohere - first on the list, the expectation that a sentence is

held true by a speaker as the result of what she takes the sentence to mean and of some belief of hers - to move from generic to detailed knowledge and understanding of the person. We can identify this first collection of assumptions with what Davidson distinguished as a special component of the Principle of Charity - the Principle of Coherence.

The second component, the Principle of Correspondence, must also come into action. It is another facet of their shared rationality that interpreter and speaker experience the world in similar ways, perceiving the same basic features, converging in congruent beliefs, moved in comparable ways by comparable motives and forces. To start uncovering specific content in the speaker's utterances and thoughts, the interpreter must try to anchor them to objects, features and events of the passing scene. As we saw, he needs the Principle of Correspondence to do that. To tentatively start translating any such observational report, the interpreter must assume that the sentence held true by the speaker is actually true, that the speaker is talking about what is going on around them, and, finally, that he himself, the interpreter, is able to recognize, and find sufficiently salient, precisely that cause and object of the speaker's report.

In Chapter 2, I considered several lines of argument supporting the validity of the Principle of Charity. Some of these arguments bore more closely on the Coherence aspect of the principle, while others made a more obvious contribution to backing up Correspondence. The investigation tried various paths, and much was added to the picture being drawn of the role of rationality and rationalization in thought, meaning and interpretation. Often, however, the inquiry stopped short of a conclusive demonstration of the Principle's adequacy.

The most successful argument, I concluded, is the argument from interpretation, and it is persuasive both with regard to Coherence and Correspondence - §2.1.3. and §2.2.2. I maintained that there is no way to interpret just one utterance or to ascribe just one thought. Each particular ascription must be integrated within a sufficiently large and coherent set of ascriptions. Furthermore, those ascriptions must be more than coherent; they must render the speaker largely correct, or true, in her beliefs - by the interpreter's own evaluation of what is correct. Too much incoherence or too much falsity compromise the identification of thoughts and meanings, and the interpreter would be unable to arrive at any content, true or false, coherent or incoherent, to ascribe to the hypothetically excessively deviant thinker and speaker. As Davidson stresses, even the recognition of error depends on a background of massive truth and coherence.

It is the same rich and vague notion of rationality that is at stake here, allowing us to understand and state that there are limits, no matter how imprecise and fluid, to the possibility of error and confusion in agents, thinkers and speakers - and, consequently, of disagreement among interlocutors. No other notion could cope with it. The prospect of making these limits, in general, more precise is an unpromising one. It is frequently up to the interpreter to judge, relative to his own expectations and overall assessment of each particular intercourse, whether he is getting something out of his interpretive efforts, whether - even if amid persistent confusion and misunderstanding - he can still make good enough sense of his interlocutor's behavior, states and utterances. Since a verdict of irrationality is very much out of the question - at least for a regular interpreter, in regular circumstances - apparent extreme disagreements must simply lead to renewed efforts of interpretation. Later in the dissertation we encountered a particularly trivial instantiation of this same dynamic, when the interpreter realizes that the speaker is not using some term in accordance with its standard or expected use and is prompted to reinterpret.

In Chapter 3, I responded to two pressing and interrelated challenges to the radical interpretability of meanings that are raised by Lepore and Ludwig in their 2005 book on Davidson.

In their first challenge, they argue that Davidson is committed to the rejection of first-person access to semantic and mental facts, and maintain that this position is both false and incoherent with other aspects of his proposals. Furthermore, the existence of a first-person access to the semantic reality compromises a tempting line of argument for the publicness, or constitutive interpretability, of meanings.

In their second challenge, they argue that the radical interpreter does not have enough resources to successfully complete his task. They maintain that a number of alternative interpretations will be equally defensible from the interpreter's position, and they take this to imply that radical interpretation, as proposed by Davidson, is not possible.

These challenges are important because radical interpretation, as Davidson defines the enterprise in its most fundamental and general features (§1.2.5), is a case, a very paradigmatic one, of interpretation via rationalization, and one which I maintain to be very much possible. However, as I have tried to make clear above (§3.4.), the apparent disagreement is not so obviously a substantive one - concerning how interpreters can and do proceed, whether or not they are able to attain interpretive success, or the role of rationalization in the process - as an

exegetical one - concerning the nature of Davidson's proposal, from its theoretical ambitions and framework to more specific details of the interpreter's method. Accordingly, instead of accepting and directly responding to the challenges, I argued that Lepore and Ludwig actually aim at a different target, not at Davidson's take on radical interpretation as I understand it.

Lepore and Ludwig's first challenge targets a certain extreme view of meaning, thought and interpretation that I dubbed *Interpretivism*. Davidson, I defended, is no interpretivist. He is in no way committed to the rejection of a first-person perspective or distinctive type of knowledge. Davidson's argument for the publicness of meaning involves not the elimination of perspectives but the requirement that there be agreement among them.

In response to Lepore and Ludwig's second challenge, I simply reiterated the strength and amplitude of scope of the Principle of Charity. This, I maintained, is enough to neutralize the immediate threat of abundant and readily demonstrable *underdetermination* of interpretation raised by Lepore and Ludwig. The purported examples of underdetermination available are promptly discredited, and the prospect of genuine ones becomes much less plausible when due attention is paid to the fact that the system of interpretation - i.e. the truth theory or meaning theory for the speaker's language - is required to deliver more than simply true theorems. The goal of interpretation is a meaning theory that renders the speaker truthful in her assertions and beliefs, but also rational, in the broad and rich sense of rationality I have been emphasizing.

In Chapter 4, I explored what I held to be some important analogies between Davidson's conceptual experiment of Radical Interpretation and the actual process and experience of interpretation from scratch carried out by children learning their first language and, more specifically, the words of their first language. In particular, after having concluded, in Chapter 2, that Davidson's recommended appeal to the Principle of Charity is a licit strategy in interpretation, in Chapter 4, I maintained that the Principle is in fact employed by children at that formative stage. For that, I turned to the scientific and empirical study of language acquisition. I presented in some detail a number of experimental studies that support the claim that word learning, in the context of first language acquisition, is best explained by attributing to children the exercise of *charitable* reconstructions of their interlocutors' thoughts and meanings, and I reported a very ample agreement on this subject among theorists in the field.

This concluded the first part of the dissertation. Rationalization was introduced, and I argued for its necessary presence on the occasions of interpretation from scratch. This was done in a markedly Davidsonian framework. In the second part, I confronted this Davidsonian general conception of meaning, language and interpretation with two alternative views on the same matters - Quine's Naturalism and Lepore and Stone's Conventionalism - views that also happen to propose forms of linguistic interpretation that do not involve rationalization.

Part II.a focused on Quine's Naturalism and is composed of Chapter 5 and 6. The mainly critical and destructive work reported in the fifth chapter prepares and explains the positive effort that is the object of the sixth.

In Chapter 5, I explained that Quine's Naturalism comprises the critical appraisal and vigorous rejection of two practices that are essential in the approach to interpretation that I favor in this dissertation, namely, the employment of mentalistic discourse and explanation and traditional and normative epistemology.

I argued that Quine fails to adequately justify his anti-mentalism both on general grounds and principles, and in local disputes over the worthiness of mentalistic explanations in particular cases. Naturalism and a commitment to science and scientific methods do not obviously or necessarily entail the repudiation of the mind or of mental postulates. Furthermore, I noted the patent success of a number of scientific disciplines that employ psychological explanations and make essential appeals to mental entities or processes. The type of investigation in word learning considered in detail in Chapter 4 is a good example, among many others. Quine's contention, or suggestion, that some of these disciplines or research programs could be rendered superfluous with the advent of more (Quinean-)naturalistic science is implausible and unsupported. In particular, as I defended in the following chapter, Quine is unsuccessful in his own attempt to inaugurate an alternative, non-mentalistic approach to language acquisition.

In response to Quine's proposed radical revision of epistemology, I maintained that the traditional and normative line of inquiry of the discipline cannot, and should not, be integrally abandoned. Although there is surely much to be praised in a naturalized line of investigation on the topics of knowledge, belief and reasoning, this does not imply that the traditional program is simply or wholly obsolete. In particular, I noted that certain epistemological questions and concerns appear to be native to our human point of view over

our own states and performances as well as others'. I also explained that there is traditional epistemology in radical interpretation, and, in general, in the interpreter's rationalizing approach to his interlocutor. The traditional, rich and vague notion of *rationality*, together with those of *evidence*, *justification*, *reason*, *making sense*, and more, cannot be missing from the enterprise.

In Chapter 6, I follow Quine in his positive efforts to elaborate an alternative view of language, meaning and interpretation - or, close enough for a relevant comparison, translation.

In the first part of the chapter, I focused on Quine's attempt to describe and start explaining our linguistic practices in (Quinean-)naturalistic terms. The discussion is centered on what Quine presents as the fundamental class of linguistic expressions - observation sentences. The keystone in Quine's semantic building is the idea that speaker's responses to these sentences are firmly and directly keyed to stimulatory circumstances. He elaborates on the idea of a (Quinean-)naturalistic similarity relation, *perceptual similarity*, and affirms it holds among stimulations that prompt like responses to observation sentences. The first stage in this Quinean enterprise is to produce behavioral criteria for the identification of observation sentences. A number of tests of observationality can be found in Quine's extensive work on the topic. I argued that two of them are well designed for Quine's purpose, and I accept Quine's criteria for the identification of observation sentences. The problem, I stressed, is that no sentence satisfies it. This is my central argument against Quine's general and sketchy proposal since, as I understand it, without observation sentences his whole schematic explanation collapses. I also expressed doubts about the naturalistic credentials of his notion of *perceptual similarity*.

In the second part of the chapter, I turned my attention to the methods of translation and interpretation that can be found, or supported, in Quine's proposals. I found no credible naturalized alternative that genuinely dispensed interpreters from rationalization, and thus no serious threat to the central claim of this dissertation.

In Part II.b, the focus shifted from interpretation from scratch to interpretation among interlocutors who already, to some extent, share a language. I endorsed an intermediary position between Davidson's radical Anti-conventionalism, and Lepore and Stone's Conventionalism. The idea of rationalization in interpretation, and the broad notion of rationality that goes with it, allowed me to delineate the dynamic conception of meaning,

language and understanding put forward there, one which recognizes the likely indispensability of a groundwork of regularities and acquired competences, but also the inexhaustible potential for impromptu change and refinement in linguistic communication.

In Chapter 7, I considered a radical form of anti-conventionalism, arguably maintained by Davidson, that affirms both the general insufficiency and the general dispensability of conventions and regularities for linguistic communication.

I concluded in favor of the general insufficiency of conventions. The reasoning behind the inference from occasional flagrant insufficiency to general, if not so evident, insufficiency, is first appealed to in this chapter: the continuous possibility of words not being used in accordance with their standard or expected meaning demands, from the interpreter, a continuous effort of rationalization.

On the other front, I maintained that the arguments against the necessity of meaning conventions are ineffective in supporting a strong and general claim of dispensability. I totally accepted that interlocutors often succeed both in understanding the linguistic utterances of others and in making their own linguistic utterances understood without sharing prior knowledge of conventions governing the expressions used. However, as I tried to show in some detail, Davidson provides us with no convincing reason to believe that improvised communication of this sort could be performed, not as an occasional and relatively exceptional feat, but as the routine. In particular, I argued that the possibility of radical interpretation in no way proves the dispensability of conventions because conventions are at the very core of the enterprise, that is, radical interpretation is all about teaching and learning prior meaning conventions, or forging new ones.

In Chapter 8, I considered Lepore and Stone's extreme form of Conventionalism, one that very strictly maintains that conventions - and other related social competences - are both sufficient and necessary for linguistic communication. Lepore and Stone claim that all linguistic meaning belongs on the conventional side of what they deem to be the fundamental divide in the uses of language - imaginative uses versus conventional uses. They claim that meaning can only arise in accordance with the conventions and prevailing standards in the community of reference. According to them, outside conventions, on the imaginative side, we only find the creative and freer uses of the language, where no precise and clear commitments to a specific meaning are actually at stake.

I carefully considered the complex and original framework that Lepore and Stone use to articulate and support these claims, one where mechanisms for social coordination take center stage, and where agents' properly meaningful engagements with the language place only very lightweight demands on their minds and lexical competence. I argued that those social mechanisms and institutions are not enough to settle or justify the ascription of conventional or standard meanings in cases where expressions are used in new or non-standard ways. In such cases, uncritical adhesion to conventions and standards simply ruins communication and mutual understanding. The decision to either adopt or abandon the conventional or expected course of interpretation can only be attained through mindreading, general interpretation and rationalization - this was the argument, already explored in the previous chapter, that allowed me to maintain that meaning conventions are generally insufficient for linguistic communication.

I also resisted Lepore and Stone's claim that conventions are not only sufficient but also always necessary for genuine linguistic coordination and communication. I argued that a full-fledged semantic status should be recognized for the non-conventional intended expression-meaning when that meaning is adequately recognized and handled by the interpreter. I noted the many crucial occasions where meanings of just this sort must be relied upon for linguistic practice, as we know it, to take place.

I have explored a particular view of human linguistic practices where rationality and rationalization figure very prominently. It is a view where the use of language is deeply integrated into the rest of our human lives and activities, where linguistic interpretation can only be endeavored in the context of a general interpretation of the person, where meanings, thoughts and actions are taken as essentially interdependent, ontologically and epistemologically speaking. This view is faithful to our natural perspective, as participants, speakers and interpreters, of the phenomenon. To be able to engage in our regular linguistic practices, to be able to make and ascribe linguistic meaning, we need the mentalistic idiom with its overtones from the normative-traditional epistemology, and we need the rational inferences and the rationalizing explanations. This is so all over, linguistic practice is never duller or less demanding. This view is firm, for instance, in rejecting any illusion of simplicity and straightforwardness in the processes involved in ostensive learning. It also reiterates the publicness of meaning, but only contingently on there being a common ground, a shared form of life, among speakers. Yet there is no threat of parochialism here, for there is

still room for all that interpreters can make sense of, or come to make sense of. Reinterpretation is a natural response to unintelligibility, and one that continuously advances language and communication. New semantic resources are forged all the time, understanding is refined, and languages are shaped as speakers go along, often in purely improvised ways. Rationally, we tend towards the repetition of successful solutions, but we are not strictly obliged to do so. The essential thing is not the regularity of the beaten path but the coordination and mutual understanding at the end of it, and we are often creative and resourceful enough to get there by other means. Finally, this view, all of this, determines a particular understanding of the interpretative task, a specific take on how to face and on what to do with our interlocutors' utterances. The solution, of course, as I have been putting it, is that linguistic interpretation requires rationalization.

References

Akhtar, Nameera, and Michael Tomasello. 1996. “Two-Year-Olds Learn Words for Absent Objects and Actions.” *British Journal of Developmental Psychology* 14 (1): 79–93. doi:10.1111/j.2044-835X.1996.tb00695.x.

———. 2000. “Social Nature of Words and Word Learning.” In *Becoming a Word Learner: A Debate on Lexical Acquisition*. Oxford University Press.

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

Armstrong, J., and E. Michaelson. 2016. “Introduction for Inquiry Symposium on Imagination and Convention.” *Inquiry* 59 (2): 139–44. doi:10.1080/0020174X.2015.1125131.

Audi, Robert. 2004. “Theoretical Rationality: Its Sources, Structure, and Scope.” In *The Oxford Handbook of Rationality*, edited by Piers Rawling and Alfred R. Mele, 17–44. Oxford: Oxford University Press.

Avramides, Anita. 1997. “Intention and Convention.” In *A Companion to the Philosophy of Language*, edited by Bob Hale and Crispin Wright. Blackwell 1997.

Baldwin, Dare A. 1991. “Infants’ Contribution to the Achievement of Joint Reference.” *Child Development* 62 (5): 875–90. doi:10.2307/1131140.

———. 1993a. “Early Referential Understanding: Infants’ Ability to Recognize Referential Acts for What They Are.” *Developmental Psychology* 29 (5): 832–43. doi:10.1037/0012-1649.29.5.832.

———. 1993b. “Infants’ Ability to Consult the Speaker for Clues to Word Reference.” *Journal of Child Language* 20 (2): 395–418. doi:10.1017/S0305000900008345.

Baldwin, Dare A., Ellen M. Markman, Brigitte Bill, Renee N. Desjardins, Jane M. Irwin, and Glynnis Tidball. 1996. “Infants’ Reliance on a Social Criterion for Establishing Word-Object Relations.” *Child Development* 67 (6): 3135–53. doi:10.2307/1131771.

Baldwin, Dare A., and Meredith Meyer. 2007. “How Inherently Social Is Language?” In *Blackwell Handbook of Language Development*, edited by Erika Hoff and Marilyn Shatz. John Wiley & Sons.

Baron-Cohen, Simon, Helen Tager-Flusberg, and Michael Lombardo, eds. 2013. *Understanding Other Minds: Perspectives from Developmental Social Neuroscience*. OUP Oxford.

Bergström, Lars. 2001. “Davidson’s Objections to Quine’s Empiricism.” In *Interpreting Davidson*, 1 edition. Stanford, Calif: CSLI Publications 2001.

Bezuidenhout, Anne. 2016. “What Properly Belongs to Grammar? A Response to Lepore and Stone.” *Inquiry* 59 (2): 175–94. doi:10.1080/0020174X.2015.1125132.

Bloom, Paul. 2000. *How Children Learn the Meanings of Words*. MIT Press.

———. 2002. “Mindreading, Communication and the Learning of Names for Things.” *Mind & Language* 17 (1-2): 37–54. doi:10.1111/1468-0017.00188.

———. 2004. “Myths of Word Learning.” In *Weaving a Lexicon*, edited by D. Geoffrey Hall and Sandra Waxman. MIT Press.

Bonjour, Laurence. 2009. *Epistemology: Classic Problems and Contemporary Responses*. 2 edition. Lanham, Md: Rowman & Littlefield Publishers.

———. 2010. “Recent Work on the Internalism–Externalism Controversy.” In *A Companion to Epistemology*, edited by Jonathan Dancy, Ernest Sosa, and Matthias Steup, 2 edition. Chichester, West Sussex, U.K ;Malden, MA: Wiley-Blackwell 2010.

Botterill, George, and Peter Carruthers. 1999. *The Philosophy of Psychology*. Cambridge, U.K. ; New York: Cambridge University Press.

Braisby, Nick, and Angus Gellatly. 2012. *Cognitive Psychology*. 2 edition. Oxford: Oxford University Press.

Burge, T. 1979. “Individualism and the Mental.” In *Foundations of Mind: Philosophical Essays, Volume 2*, 1 edition, 100–150. Oxford : Oxford ; New York: Clarendon Press 2007.

———. 2006. “Postscript to ‘Individualism and the Mental.’” In *Foundations of Mind: Philosophical Essays, Volume 2*, 1 edition, 151–81. Oxford : Oxford ; New York: Clarendon Press 2007.

———. 2007. *Foundations of Mind: Philosophical Essays, Volume 2*. 1 edition. Oxford : Oxford ; New York: Clarendon Press.

Burge, T., Donald. 1999. “Comprehension and Interpretation.” In *The Library of Living Philosophers: Donald Davidson*, edited by F. Hahn. Open Court 1999.

Burge, Tyler. 2010. *Origins of Objectivity*. 1 edition. Oxford ; New York: Oxford University Press.

Carnap, Rudolf. 1928. *The Logical Structure of the World and Pseudoproblems in Philosophy*. Translated by Rolf A. George. Berkeley: University of California Press 1967.

Caro, Mario De, ed. 1999. *Interpretations and Causes: New Perspectives on Donald Davidson’s Philosophy*. Kluwer.

Carston, Robyn. 2016. “Linguistic Conventions and the Role of Pragmatics.” *Mind & Language* 31 (5): 612–24. doi:10.1111/mila.12122.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. The MIT Press.

———. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. Greenwood Publishing Group.

———. 1992. “Language and Interpretation: Philosophical Reflections and Empirical Inquiry.” In *New Horizons in the Study of Language and Mind*. Cambridge University Press 2000.

———. 1994. “Naturalism and Dualism in the Study of Language and Mind.” In *New Horizons in the Study of Language and Mind*. Cambridge University Press 2000.

———. 1995. “Language as a Natural Object.” In *New Horizons in the Study of Language and Mind*. Cambridge University Press 2000.

———. 2000a. “Language from an Internalist Perspective.” In *New Horizons in the Study of Language and Mind*. Cambridge University Press 2000.

———. 2000b. *New Horizons in the Study of Language and Mind*. Cambridge University Press.

———. 2009. “Conclusion.” In *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*. OUP Oxford.

Chomsky, Noam, and James McGilvray. 2012. *The Science of Language: Interviews with James McGilvray*. Cambridge University Press.

Cohen, L. Jonathan. 1981. “Can Human Irrationality Be Experimentally Demonstrated?” *Behavioral and Brain Sciences* 4 (3): 317–70.

Collins, John. 2016. “Putting Syntax First: On Convention and Implicature in Imagination and Convention.” *Mind & Language* 31 (5): 635–45. doi:10.1111/mila.12124.

Dancy, Jonathan, Ernest Sosa, and Matthias Steup, eds. 2010. *A Companion to Epistemology*. 2 edition. Chichester, West Sussex, U.K ;Malden, MA: Wiley-Blackwell.

Davidson, Donald. 1963. “Actions, Reasons, and Causes.” In *Essays on Actions and Events*. Oxford u.a.: Clarendon Press 2001.

———. 1965. “Theories of Meaning and Learnable Languages.” In *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984.

———. 1967. “Truth and Meaning.” In *Inquiries into Truth and Interpretation*, 17–36. Oxford: Clarendon Press 1984. Originally published in Synthèse 17(1967):304–23.

———. 1969. “On Saying That.” In *Inquiries into Truth and Interpretation*, 130–46. Oxford: Clarendon Press 1984. Originally published in Synthèse 19(1969):130–46.

———. 1970. “Mental Events.” In *Inquiries into Truth and Interpretation*, 207–24. Oxford: Clarendon Press 1984. Originally published in Foster, L., and Swanson, J.W. (eds). Experience and Theory. Amherst: The University of Massachusetts Press (1970).

———. 1973. “Radical Interpretation.” In *Inquiries into Truth and Interpretation*, 125–39. Oxford: Clarendon Press 1984. Originally published in Dialectica 27(1973):313–328.

———. 1974a. “Belief and the Basis of Meaning.” In *Inquiries into Truth and Interpretation*, 141–54. Oxford: Clarendon Press 1984. Originally published in Synthèse 27(1974):329–343.

———. 1974b. “On the Very Idea of a Conceptual Scheme.” In *Inquiries into Truth and Interpretation*, 183–98. Oxford: Clarendon Press 1984. Originally published in Proceedings and Addresses of the American Philosophical Association, Volume 47 (1974).

———. 1974c. “Psychology as Philosophy.” In *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984.

———. 1975. “Thought and Talk.” In *Inquiries into Truth and Interpretation*, 155–70. Oxford: Clarendon Press 1984. Originally published in Guttenplan, S. (ed.). Mind and Language. Oxford: Oxford University Press (1975).

———. 1976a. “Hempel on Explaining Action.” In *Essays on Actions and Events*. Oxford u.a.: Clarendon Press 2001.

———. 1976b. “Reply to Foster.” In *Inquiries into Truth and Interpretation*, 171–79. Oxford: Clarendon Press 1984. Originally published in Evans, G. and McDowell, J. (eds). Truth and Meaning. Essays in Semantics. Oxford: Oxford University Press (1976), pp. 33–41.

———. 1977a. “Reality without Reference.” In *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984.

———. 1977b. “The Method of Truth in Metaphysics.” In *Inquiries into Truth and Interpretation*, 199–214. Oxford: Clarendon Press 1984.

———. 1979. “The Inscrutability of Reference.” In *Inquiries into Truth and Interpretation*, 199–214. Oxford: Clarendon Press 1984.

———. 1980. “A Unified Theory of Thought, Meaning, and Action.” In *Problems of Rationality*. Oxford: Clarendon Press 2004.

———. 1982a. “Communication and Convention.” In *Inquiries into Truth and Interpretation*, 199–214. Oxford: Clarendon Press 1984.

———. 1982b. “Communication and Convention.” In *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984.

———. 1982c. “Empirical Content.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1982d. “Rational Animals.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1983. “A Coherence Theory of Truth and Knowledge.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1984a. “Expressing Evaluations.” In *Problems of Rationality*. Oxford: Clarendon Press 2004.

———. 1984b. “First Person Authority.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1984c. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

———. 1984d. “Introduction to Inquiries into Truth and Interpretation.” In *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984.

———. 1985a. “A New Basis for Decision Theory.” *Theory and Decision* 18 (1): 87–98.

———. 1985b. “Reply to Smart.” In *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, edited by Ernest Lepore and Brian P. McLaughlin. B. Blackwell 1985.

———. 1986. “A Nice Derangement of Epitaphs.” In *Truth, Language, and History*. Oxford: Clarendon Press 2005.

———. 1987. “Knowing One’s Own Mind.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1988. “Epistemology and Truth.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1989. “James Joyce and Humpty Dumpty.” In *Truth, Language, and History*. Oxford: Clarendon Press 2005.

———. 1990a. “Epistemology Externalized.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1990b. “Meaning, Truth and Evidence.” In *Truth, Language, and History*, 47–62. Oxford: Clarendon Press 2005. Originally published in R. Barrett and R. Gibson, eds. (1990). Perspectives on Quine . Oxford: Basil Blackwell, pp. 68–79.

———. 1990c. “Representation and Interpretation.” In *Problems of Rationality*. Oxford: Clarendon Press 2004.

———. 1990d. “The Structure and Content of Truth.” *Journal of Philosophy* 87 (6): 279–328.

———. 1990e. “Turing’s Test.” In *Problems of Rationality*. Oxford: Clarendon Press 2004.

———. 1991. “Three Varieties of Knowledge.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1992a. “The Emergence of Thought.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1992b. “The Second Person.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1992c. “The Third Man.” In *Truth, Language, and History*. Oxford: Clarendon Press 2005.

———. 1993. “Locating Literary Language.” In *Truth, Language, and History*. Oxford: Clarendon Press 2005.

———. 1994a. “Radical Interpretation Interpreted.” *Philosophical Perspectives* 8: 121–28.

———. 1994b. “The Social Aspect of Language.” In *Truth, Language, and History*. Oxford: Clarendon Press 2005.

———. 1995a. “Could There Be a Science of Rationality?” In *Problems of Rationality*. Oxford: Clarendon Press 2004.

———. 1995b. “The Objectivity of Values.” In *Problems of Rationality*. Oxford: Clarendon Press 2004.

———. 1995c. “The Problem of Objectivity.” In *Problems of Rationality*. Oxford: Clarendon Press 2004.

———. 1997a. “Indeterminism and Antirealism.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1997b. “Seeing Through Language.” In *Truth, Language, and History*. Oxford: Clarendon Press 2005.

———. 1998. “The Irreducibility of the Concept of the Self.” In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001.

———. 1999a. “Interpretation: Hard in Theory, Easy in Practice.” In *Interpretations and Causes: New Perspectives on Donald Davidson’s Philosophy*, edited by Mario De Caro. Kluwer.

———. 1999b. “Reply to Cutrofello.” In *The Library of Living Philosophers: Donald Davidson*, edited by F. Hahn. Open Court 1999.

———. 1999c. “Reply to Stroud.” In *The Library of Living Philosophers: Donald Davidson*, edited by F. Hahn. Open Court 1999.

———. 2001a. “Comments on Karlovy Papers.” In *Interpreting Davidson*, edited by Petr Kotatko, Peter Pagin, and Gabriel Segal, 1 edition. Stanford, Calif: CSLI Publications 2001.

———. 2001b. *Essays on Actions and Events*. 2 edition. Oxford u.a.: Clarendon Press.

———. 2001c. “Externalisms.” In *Interpreting Davidson*, edited by Petr Kotatko, Peter Pagin, and Gabriel Segal, 1 edition. Stanford, Calif: CSLI Publications 2001.

———. 2001d. *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press.

———. 2001e. “What Thought Requires.” In *Problems of Rationality*. Oxford: Clarendon Press 2004.

———. 2004. *Problems of Rationality*. Oxford: Clarendon Press.

———. 2005a. *Truth, Language, and History*. Oxford: Clarendon Press.

———. 2005b. *Truth & Predication*. Cambridge, Mass. u.a.: Belknap Press of Harvard University Press.

Davidson, Donald, and Ernest Lepore. 2004. “An Interview with Donald Davidson.” In *Problems of Rationality*. Oxford: Clarendon Press 2004.

Demopolous, William. 2007. “Carnap on the Rational Reconstruction of Scientific Theories.” In *The Cambridge Companion to Carnap*, edited by Michael Friedman and Richard Creath. Cambridge, UK ; New York: Cambridge University Press 2007.

Dever, Josh. 2012. “Formal Semantics.” In *The Continuum Companion to the Philosophy of Language*, edited by Manuel Garcia-Carpintero and Max Kolbel, annotated edition edition. London u.a.: Bloomsbury Academic.

Diesendruck, Gil. 2007. “Mechanisms of Word Learning.” In *Blackwell Handbook of Language Development*, edited by Erika Hoff and Marilyn Shatz. John Wiley & Sons.

Diesendruck, G., and L. Markson. 2001. “Children’s Avoidance of Lexical Overlap: A Pragmatic Account.” *Developmental Psychology* 37 (5): 630–41.

Ditmarsch, Hans van, and Barteld Kooi. 2015. *One Hundred Prisoners and a Light Bulb*. Springer.

Elio, Renée, ed. 2002. *Common Sense, Reasoning, & Rationality*. Oxford University Press.

Fodor, Jerry, and Ernest Lepore. 1992. *Holism: A Shopper’s Guide*. 1 edition. Oxford u.a.: Wiley-Blackwell.

Fodor, Jerry, and Ernie Lepore. 1994. “Is Radical Interpretation Possible?” *Philosophical Perspectives* 8: 101–19. doi:10.2307/2214165.

Foley, Hugh, and Margaret Matlin. 2009. *Sensation and Perception*. 5 edition. Boston: Psychology Press.

Føllesdal, Dagfinn. 2014. “Developments in Quine’s Behaviorism.” In *A Companion to W. V. O. Quine*, edited by Gilbert Harman and Ernest Lepore, 1 edition. Hoboken, New Jersey: Wiley-Blackwell 2014.

Friedman, Michael. 2007. “The Aufbau and the Rejection of Methaphysics.” In *The Cambridge Companion to Carnap*, edited by Michael Friedman and Richard Creath. Cambridge, UK ; New York: Cambridge University Press 2007.

Friedman, Michael, and Richard Creath, eds. 2007. *The Cambridge Companion to Carnap*. Cambridge, UK ; New York: Cambridge University Press.

Gallaway, Clare, and Brian J. Richards, eds. 1994. *Input and Interaction in Language Acquisition*. Cambridge University Press.

García-Carpintero, Manuel. 2012. “Foundational Semantics I: Descriptive Accounts.” *Philosophy Compass* 7 (6): 397–409.

Garcia-Carpintero, Manuel, and Max Kolbel, eds. 2012. *The Continuum Companion to the Philosophy of Language*. Annotated edition edition. London u.a.: Bloomsbury Academic.

Gibson, James J. 1979. *The Ecological Approach to Visual Perception: Classic Edition*. 1 edition. New York, London: Psychology Press.

Gjelsvik, Olav. 2014. “Quine on Observationality.” In *A Companion to W. V. O. Quine*, edited by Gilbert Harman and Ernest Lepore, 1 edition. Hoboken, New Jersey: Wiley-Blackwell 2014.

Gleitman, Lila. 2009. “The Learned Component of Language Learning.” In *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*. OUP Oxford.

Gleitman, Lila, and Cyntia Fisher. 2005. “Universal Aspects of Word Learning.” In *The Cambridge Companion to Chomsky*, edited by James McGilvray. Cambridge University Press.

Glüer, Kathrin. 2006. “The Status of Charity I: Conceptual Truth or a Posteriori Necessity?” *International Journal of Philosophical Studies* 14 (3): 337–59.

———. 2011. *Donald Davidson: A Short Introduction*. 1 edition. New York: Oxford University Press.

———. 2012. “Theories of Meaning and Truth Conditions.” In *The Continuum Companion to the Philosophy of Language*, edited by Manuel Garcia-Carpintero and Max Kolbel. London: Bloomsbury Academic.

———. 2013. “Convention and Meaning.” In *A Companion to Donald Davidson*, edited by Ernest Lepore and Kirk Ludwig, 1 edition. Malden, Mass. u.a.: Wiley-Blackwell 2013.

Gogol, Nikolai. 2016. “Memoirs of a Madman.” In *The Mantle and Other Stories, by Nicholas Gogol—A Project Gutenberg eBook*, translated by Claud Field, http://www.gutenberg.org/files/36238/36238-h/36238-h.htm#Page_107 ed. Accessed November 15. http://www.gutenberg.org/files/36238/36238-h/36238-h.htm#Page_107.

Goldman, Alvin I. 2006. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.

Golinkoff, Roberta Michnick, Kathryn Hirsh-Pasek, Lois Bloom, Linda B. Smith, Amanda L. Woodward, Nameera Akhtar, Michael Tomasello, and George Hollich. 2000. *Becoming a Word Learner: A Debate on Lexical Acquisition*. Oxford University Press.

Goodman, Nelson. 1954. *Fact, Fiction, and Forecast*. Harvard University Press 1983.

Gregory, Paul. 2008. *Quine’s Naturalism: Language, Theory, and the Knowing Subject*. Continuum.

Grice, P. 1957. “Meaning.” In *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press 1989.

———. 1968. “Utterer’s Meaning, Sentence Meaning, and Word-Meaning.” In *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press 1989.

———. 1969. “Utterer’s Meaning and Intentions.” In *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press 1989.

———. 1982. “Meaning Revisited.” In *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press 1989.

———. 1989. *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press.

Gross, Steven. 2012. “Davidson, First-Person Authority, and the Evidence for Semantics.” In *Donald Davidson on Truth, Meaning, and the Mental*, edited by Gerhard Preyer. Oxford University Press 2012.

Gunnar, Megan R., and Michael Maratsos, eds. 1992. *Modularity and Constraints in Language and Cognition*. Psychology Press.

Guttenplan, Samuel, ed. 1975. *Mind and Language*. Oxford: Oxford University Press.

Hahn, F., ed. 1999. *The Library of Living Philosophers: Donald Davidson*. Open Court.

Hahn, Lewis Edwin, and Paul Arthur Schilpp, eds. 1986. *The Philosophy of W. V. Quine*. 2nd edition 1998. La Salle, Ill: Open Court.

Hale, Bob, and Crispin Wright. 1997. *A Companion to the Philosophy of Language*. Wiley-Blackwell.

Hale, Bob, Crispin Wright, and Alexander Miller, eds. 2017. *A Companion to the Philosophy of Language*. 2nd Revised edition edition. Chichester, West Sussex, UK ; Malden, MA: Wiley-Blackwell.

Hall, D. Geoffrey, and Sandra Waxman, eds. 2004. *Weaving a Lexicon*. MIT Press.

Hardcastle, Valerie. 2009. "The Interface between Psychology and Neuroscience." In *The Routledge Companion to Philosophy of Psychology*, edited by John Symons and Paco Calvo. London ; New York: Routledge 2009.

Harman, Gilbert. 2014. "Indeterminacy, Relativity, and Behaviorism." In *A Companion to W. V. O. Quine*, edited by Gilbert Harman and Ernest Lepore, 1 edition. Hoboken, New Jersey: Wiley-Blackwell 2014.

Harman, Gilbert, and Ernest Lepore, eds. 2014. *A Companion to W. V. O. Quine*. 1 edition. Hoboken, New Jersey: Wiley-Blackwell.

Hawthorne, John. 2016. "Some Remarks on Imagination and Convention." *Mind & Language* 31 (5): 625–34. doi:10.1111/mila.12123.

Hirsh-Pasek, Kathryn, Roberta Michnick Golinkoff, E. Hennon, and Mandy Maguire. 2004. "Hybrid Theories at the Frontier of Developmental Psychology: The Emergentist Coalition Model of Word Learning as a Case in Point." In *Weaving a Lexicon*, edited by D. Geoffrey Hall and Sandra Waxman. MIT Press.

Hoff, Erika, and Marilyn Shatz, eds. 2007. *Blackwell Handbook of Language Development*. John Wiley & Sons.

Hollich, George J., Kathy Hirsh-Pasek, Roberta Michnick Golinkoff, Rebecca J. Brand, Ellie Brown, He Len Chung, Elizabeth Hennon, Camille Rocroi, and Lois Bloom. 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 – 135.

Horn, Laurence R. 2016. "Conventional Wisdom Reconsidered." *Inquiry* 59 (2): 145–62. doi:10.1080/0020174X.2015.1125133.

Hylton, Peter. 2007. *Quine*. 1 edition. New York: Routledge.

———. 2014. "Quine's Naturalism Revisited." In *A Companion to W. V. O. Quine*, edited by Gilbert Harman and Ernest Lepore, 1 edition. Hoboken, New Jersey: Wiley-Blackwell 2014.

Jackman, Henry. 2003. "Charity, Self-Interpretation, and Belief." *Journal of Philosophical Research* 28: 143–68.

Johnson-Laird, P. N., and P. C. Wason. 1977a. "A Theoretical Analysis of Insight into a Reasoning Task, and Postscript." In *Thinking: Readings in Cognitive Science*. Cambridge University Press.

———. , eds. 1977b. *Thinking: Readings in Cognitive Science*. Cambridge University Press.

Kelly, Thomas. 2014. "Quine and Epistemology." In *A Companion to W. V. O. Quine*, edited by Gilbert Harman and Ernest Lepore, 1 edition. Hoboken, New Jersey: Wiley-Blackwell 2014.

Kemp, Gary. 2012. *Quine versus Davidson: Truth, Reference, and Meaning*. 1 edition. Oxford : New York: Oxford University Press.

Kim, Jaegwon. 1988. “What Is ‘Naturalized Epistemology?.’” *Philosophical Perspectives* 2 (January): 381–405. doi:10.2307/2214082.

Korta, Kepa, E. Sosa, and Xabier Arrazola, eds. 1999. *Cognition, Agency and Rationality: Proceedings of the Fifth International Colloquium on Cognitive Science*. Springer Science & Business Media.

Kotatko, Petr, Peter Pagin, and Gabriel Segal, eds. 2001. *Interpreting Davidson*. 1 edition. Stanford, Calif: CSLI Publications.

Kripke, S. A. 1972. *Naming and Necessity*. (1980) ed. Harvard University Press.

———. 1977. “Speaker’s Reference and Semantic Reference.” In *Philosophical Troubles: Collected Papers, Volume 1*, 1 edition. New York: Oxford University Press 2011.

———. 1979. “A Puzzle about Belief.” In *Philosophical Troubles: Collected Papers, Volume 1*, 1 edition. New York: Oxford University Press 2011.

———. 1982. *Wittgenstein on Rules and Private Language*. Harvard University Press.

———. 2011. *Philosophical Troubles: Collected Papers, Volume 1*. 1 edition. New York: Oxford University Press.

Lepore, Ernest, and Kirk Ludwig. 2007. *Donald Davidson’s Truth-Theoretic Semantics*. Clarendon Press.

———. 2012. “Introduction: Davidson’s Philosophical Project.” In *Donald Davidson on Truth, Meaning, and the Mental*, edited by Gerhard Preyer. Oxford University Press 2012.

———. , eds. 2013. *A Companion to Donald Davidson*. 1 edition. Malden, Mass. u.a.: Wiley-Blackwell.

Lepore, Ernest, and Brian P. McLaughlin, eds. 1985. *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. B. Blackwell.

Lepore, Ernie, and Kirk Ludwig. 2005. *Donald Davidson: Meaning, Truth, Language, and Reality*. Oxford University Press Uk.

Lepore, Ernie, and Barry C. Smith, eds. 2006. *The Oxford Handbook of Philosophy of Language*. 1 edition. Oxford u.a.: Oxford University Press.

Lepore, Ernie, and Matthew Stone. 2015. *Imagination and Convention: Distinguishing Grammar and Inference in Language*. 1 edition. Oxford: Oxford University Press.

———. 2016a. “The Breadth of Semantics: Reply to Critics.” *Inquiry* 59 (2): 195–206. doi:10.1080/0020174X.2015.1125134.

———. 2016b. “Semantics, Coherence, and Intentions: Reply to Carston, Collins and Hawthorne.” *Mind & Language* 31 (5): 646–54. doi:10.1111/mila.12125.

Lewis, D. 1969. *Convention: A Philosophical Study*. (2002) ed. Oxford: Blackwell.

———. 1975. “Language and Languages.” In *Philosophical Papers: Volume I*, 1 edition, 163–88. New York: Oxford University Press.

———. 1979. “Scorekeeping in a Language Game.” In *Philosophical Papers: Volume I*, 1 edition, 233–49. New York: Oxford University Press 1983.

———. 1983. *Philosophical Papers: Volume I*. 1 edition. New York: Oxford University Press.

Lewis, David. 1970. “General Semantics.” *Synthese* 22 (1-2): 18–67.

Lieven, Elena V. M. 1994. “Crosslinguistic and Crosscultural Aspects of Language Addressed to Children.” In *Input and Interaction in Language Acquisition*. Cambridge University Press.

Liszkowski, Ulf, Malinda Carpenter, Tricia Striano, and Michael Tomasello. 2006. “12- and 18-Month-Olds Point to Provide Information for Others.” *Journal of Cognition and Development* 7 (2): 173–87. doi:10.1207/s15327647jcd0702_2.

Ludlow, Peter. 2014. *Living Words: Meaning Underdetermination and the Dynamic Lexicon*. Oxford University Press.

Ludwig, Kirk. 2004. “Rationality, Language and the Principle of Charity.” In *The Oxford Handbook of Rationality*, edited by Alfred R. Mele and Piers Rawling, 1 edition. Oxford ; New York: Oxford University Press 2004.

Manning, Richard N. 2012. “Taking Back the Excitement: Construing ‘theoretical Concepts’ so as to Avoid the Threat of Underdetermination.” In *Donald Davidson on Truth, Meaning, and the Mental*, edited by Gerhard Preyer. Oxford University Press 2012.

Markman, Ellen M. 1989. *Categorization and Naming in Children: Problems of Induction*. MIT Press.

———. 1992. “Constraints on Word Learning: Speculations about Their Nature, Origins, and Domain Specificity.” In *Modularity and Constraints in Language and Cognition*. Psychology Press.

McCauley, Robert. 2007. “Reduction: Models of Cross-Scientific Relations and Their Implications for the Psychology-Neuroscience Interface.” In *Philosophy of Psychology and Cognitive Science*, edited by Paul Thagard, Dov M. Gabbay, and John Woods. Amsterdam: North Holland 2007.

McGilvray, James. 2005a. “Introduction.” In *The Cambridge Companion to Chomsky*, edited by James McGilvray. Cambridge University Press.

———. , ed. 2005b. *The Cambridge Companion to Chomsky*. Cambridge University Press.

Mele, Alfred R., and Piers Rawling, eds. 2004. *The Oxford Handbook of Rationality*. 1 edition. Oxford ; New York: Oxford University Press.

Moshman, David. 2014. *Epistemic Cognition and Development: The Psychology of Justification and Truth*. Psychology Press.

Myers, Robert H., and Claudine Verheggen. 2016. *Donald Davidson’s Triangulation Argument: A Philosophical Inquiry*. Routledge.

Neale, Stephen. 1992. “Paul Grice and the Philosophy of Language.” *Linguistics and Philosophy* 15 (5): 509–59.

Niiniluoto, I., Matti Sintonen, and Jan Wolenski, eds. 2004. *Handbook of Epistemology*. Springer Science & Business Media.

Nisbett, Richard E., and Lee Ross. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. Prentice-Hall.

Nolan, Daniel. 2005. *David Lewis*. Chesham: Routledge.

Orenstein, A., and Petr Kotatko, eds. 2000. *Knowledge, Language and Logic: Questions for Quine*. Kluwer Academic Print on Demand.

Pagin, Peter. 2006a. "Meaning Holism." In *The Oxford Handbook of Philosophy of Language*, edited by Ernest Lepore and Barry C. Smith. OUP Oxford 2006.

———. 2006b. "The Status of Charity II: Charity, Probability, and Simplicity." *International Journal of Philosophical Studies* 14 (3): 361–83.

———. 2008. "What Is Communicative Success?" *Canadian Journal of Philosophy* 38 (1): 85–115. doi:10.1353/cjp.0.0011.

———. 2013. "Radical Interpretation and the Principle of Charity." In *A Companion to Donald Davidson*, edited by Ernest Lepore and Kirk Ludwig, 1 edition. Malden, Mass. u.a.: Wiley-Blackwell 2013.

Piattelli-Palmarini, Massimo, Juan Uriagereka, and Pello Salaburu, eds. 2009. *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*. OUP Oxford.

Pietroski, Paul. 1994. "A Defense of Derangement." *Canadian Journal of Philosophy* 24 (1): 95–117.

Pinker, Steven. 1994. *The Language Instinct*. New York: Harper Collins.

Preyer, Gerhard. 2012. *Donald Davidson on Truth, Meaning, and the Mental*. Oxford University Press.

Psillos, Stathis, and Martin Curd, eds. 2013. *The Routledge Companion to Philosophy of Science*. 2 edition. London ; New York: Routledge.

Putnam, Hilary. 1975. "The Meaning of 'Meaning.'" In *Philosophical Papers, Vol. 2: Mind, Language and Reality*, 215–71. Cambridge University Press 1979.

———. 1979. *Philosophical Papers, Vol. 2: Mind, Language and Reality*. Cambridge University Press.

———. 1982. "Why Reason Can't Be Naturalized." *Synthese* 52 (1): 3–23. doi:10.1007/BF00485252.

———. 1988. *Representation and Reality*. Revised ed. edition 1991. Cambridge, Mass.: A Bradford Book.

———. 1992. "Why Functionalism Didn't Work?" In *Words and Life*. Harvard University Press 1992.

Putnam, Hilary, and James Conant. 1995. *Words and Life*. Harvard University Press.

Quine, Willard V. 1974. *The Roots of Reference*. La Salle, Ill.: Open Court Publishing Company.

Quine, Willard Van Orman. 1951. "Two Dogmas of Empiricism." In *From a Logical Point of View*. Cambridge, Mass: Harvard University Press.

———. 1953. *From a Logical Point of View*. Revised edition 1980. Cambridge, Mass: Harvard University Press.

———. 1957. “The Scope and Language of Science.” In *The Ways of Paradox and Other Essays*. Cambridge, Mass.: Harvard University Press 1966.

———. 1960. *Word and Object*. New edition, 2013. The MIT Press.

———. 1966. *The Ways of Paradox and Other Essays*. Revised edition 1976. Cambridge, Mass.: Harvard University Press.

———. 1969a. “Epistemology Naturalized.” In *Ontological Relativity & Other Essays*, First Printing edition. New York: Columbia University Press 1969.

———. 1969b. *Ontological Relativity & Other Essays*. First Printing edition. New York: Columbia University Press.

———. 1975. “Mind and Verbal Dispositions.” In *Mind and Language*, edited by Samuel Guttenplan. Oxford: Oxford University Press 1975.

———. 1978. “Facts of the Matter.” *Southwestern Journal of Philosophy* 9 (2): 155–69.

———. 1981a. “Empirical Content.” In *Theories and Things*. Cambridge, Mass.: Belknap Press.

———. 1981b. “Responses.” In *Theories and Things*. Cambridge, Mass.: Belknap Press.

———. 1981c. *Theories and Things*. Cambridge, Mass.: Belknap Press.

———. 1986a. “Reply to Putnam.” In *The Philosophy of W. V. Quine*, edited by Lewis Edwin Hahn and Paul Arthur Schilpp, 2nd edition 1998. La Salle, Ill: Open Court.

———. 1986b. “Reply to White.” In *The Philosophy of W. V. Quine*, edited by Lewis Edwin Hahn and Paul Arthur Schilpp, 2nd edition 1998. La Salle, Ill: Open Court.

———. 1987. *Quiddities: An Intermittently Philosophical Dictionary*. Cambridge, Mass: Belknap Press.

———. 1990. *Pursuit of Truth*. Harvard University Press.

———. 1995. *From Stimulus to Science*. Harvard UP.

———. 2000a. “I, You, and It: An Epistemological Triangle.” In *Knowledge, Language and Logic: Questions for Quine*, edited by A. Orenstein and Petr Kotatko. Kluwer Academic Print on Demand.

———. 2000b. “Response to George.” In *Knowledge, Language and Logic: Questions for Quine*, edited by A. Orenstein and Petr Kotatko. Kluwer Academic Print on Demand.

Quine, W. V. 1970. “Methodological Reflections on Current Linguistic Theory.” *Synthese* 21 (3/4): 386–98.

———. 1993. “In Praise of Observation Sentences.” *The Journal of Philosophy* 90 (3): 107–16. doi:10.2307/2940954.

———. 1995. “Naturalism; Or, Living Within One’s Means.” *Dialectica* 49 (2/4): 251–61.

———. 1996. “Progress on Two Fronts.” *The Journal of Philosophy* 93 (4): 159–63. doi:10.2307/2940885.

Ramberg, B. T. (1989). *Donald Davidson’s Philosophy of Language: An Introduction*. Blackwell.

Ramsey, William. 2016. “Eliminative Materialism.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/materialism-eliminative/.

Rawling, Piers. 2013. “Davidson’s Measurement-Theoretic Analogy.” In *A Companion to Donald Davidson*, edited by Ernest Lepore and Kirk Ludwig, 1 edition, 247–63. Malden, Mass. u.a.: Wiley-Blackwell 2013.

Rysiew, Patrick. 2008. “Rationality Disputes – Psychology and Epistemology.” *Philosophy Compass* 3 (6): 1153–76. doi:10.1111/j.1747-9991.2008.00178.x.

———. 2017. “Naturalism in Epistemology.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2017. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2017/entries/epistemology-naturalized/.

Samuels, Richard, Stephen P. Stich, and Michael Bishop. 2002. “Ending the Rationality Wars: How to Make DIsputes about Human Rationality Disappear.” In *Common Sense, Reasoning, & Rationality*, edited by Renée Elio. Oxford University Press.

Samuels, Richard, Stephen P. Stich, and Luc Faucher. 2004. “Reason and Rationality.” In *Handbook of Epistemology*, edited by I. Niiniluoto, Matti Sintonen, and Jan Wolenski. Springer Science & Business Media 2004.

Schiffer, Stephen. 2017. “Intention and Convention in the Theory of Meaning.” In *A Companion to the Philosophy of Language*, edited by Bob Hale, Crispin Wright, and Alexander Miller, 2nd Revised edition edition. Chichester, West Sussex, UK ; Malden, MA: Wiley-Blackwell 2017.

Searle, John R. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press.

Sinclair, Robert. 2014. “Quine on Evidence.” In *A Companion to W. V. O. Quine*, edited by Gilbert Harman and Ernest Lepore, 1 edition. Hoboken, New Jersey: Wiley-Blackwell 2014.

Smith, Barry C. 2006. “What I Know When I Know a Language.” In *The Oxford Handbook of Philosophy of Language*, edited by Ernest Lepore and Barry C. Smith. Oxford University Press 2006.

Smith, Linda B. 2000a. “Avoiding Associations When It’s Behaviorism You Really Hate.” In *Becoming a Word Learner: A Debate on Lexical Acquisition*. Oxford University Press.

———. 2000b. “Learning How to Learn WordsAn Associative Crane.” In *Becoming a Word Learner: A Debate on Lexical Acquisition*. Oxford University Press.

Smullyan, Raymond M. 1978. *What Is the Name of This Book?: The Riddle of Dracula and Other Logical Puzzles*. Unabridged edition. Prentice-Hall, Englewood Cliffs, N.J.

Sosa, E. 1999. “Are Humans Rational?” In *Cognition, Agency and Rationality: Proceedings of the Fifth International Colloquium on Cognitive Science*, edited by Kepa Korta, E. Sosa, and Xabier Arrazola. Springer Science & Business Media.

Southgate, Victoria. 2013. “Early Manifestations of Mindreading.” In *Understanding Other Minds: Perspectives from Developmental Social Neuroscience*, edited by Simon Baron-Cohen, Helen Tager-Flusberg, and Michael Lombardo. OUP Oxford.

Stalnaker, Robert C. 1997. “Reference and Necessity.” In *A Companion to the Philosophy of Language*, edited by Bob Hale and Crispin Wright. Blackwell.

Stein, Edward. 1996. *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science*. 1 edition. Oxford : Oxford ; New York: Clarendon Press.

Stich, Stephen P. 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief*. MIT Press.

———. 1993. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. New edition edition. Cambridge, Mass.: MIT Press.

Symons, John, and Paco Calvo, eds. 2009. *The Routledge Companion to Philosophy of Psychology*. London ; New York: Routledge.

Szabó, Zoltán Gendler. 2016. “In Defense of Indirect Communication.” *Inquiry* 59 (2): 163–74. doi:10.1080/0020174X.2015.1125135.

Tarski, Alfred. 1933. “The Concept of Truth in Formalized Languages.” In *Logic, Semantics, Metamathematics*, translated by J. H. Woodger. Oxford University Press 1956.

———. 1944. “The Semantic Conception of Truth: And the Foundations of Semantics.” *Philosophy and Phenomenological Research* 4 (3): 341–76.

Thagard, Paul. 2013. “Cognitive Science.” In *The Routledge Companion to Philosophy of Science*, edited by Stathis Psillos and Martin Curd, 2 edition. London ; New York: Routledge.

Thagard, Paul, Dov M. Gabbay, and John Woods, eds. 2007. *Philosophy of Psychology and Cognitive Science*. Amsterdam: North Holland.

Tomasello, Michael. 2003. *Constructing a Language*. Harvard University Press.

Tomasello, Michael, and Nameera Akhtar. 1995. “Two-Year-Olds Use Pragmatic Cues to Differentiate Reference to Objects and Actions.” *Cognitive Development* 10 (2): 201–24. doi:10.1016/0885-2014(95)90009-8.

Tomasello, Michael, and Michelle E. Barton. 1994. “Learning Words in Nonostensive Contexts.” *Developmental Psychology* 30 (5): 639–50. doi:10.1037/0012-1649.30.5.639.

Tomasello, Michael, Randi Strosberg, and Nameera Akhtar. 1996. “Eighteen-Month-Old Children Learn Words in Non-Ostensive Contexts.” *Journal of Child Language* 23 (1): 157–76. doi:10.1017/S0305000900010138.

Tversky, Amos, and Daniel Kahneman. 1983. “Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment.” *Psychological Review* 90 (4): 293–315. doi:10.1037/0033-295X.90.4.293.

Uebel, Thomas. 2007. “Carnap and the Vienna Circle: Rational Reconstructionism Refined.” In *The Cambridge Companion to Carnap*, edited by Michael Friedman and Richard Creath. Cambridge, UK ; New York: Cambridge University Press 2007.

Verheggen, Claudine. 2013. “Triangulation.” In *A Companion to Donald Davidson*, edited by Ernest Lepore and Kirk Ludwig, 1 edition, 456–71. Malden, Mass. u.a.: Wiley-Blackwell 2013.

Wier, Alan. 2014. “Quine’s Naturalism.” In *A Companion to W. V. O. Quine*, edited by Gilbert Harman and Ernest Lepore, 1 edition. Hoboken, New Jersey: Wiley-Blackwell 2014.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Edited by Hacker and Schulte. 4th Edition. Wiley-Blackwell. 2009

Wrenn, Chase. 2006. “Epistemology as Engineering?” *Theoria* 72 (1): 60–79.

Yu, Chen, and Linda B. Smith. 2007. “Rapid Word Learning under Uncertainty via Cross-Situational Statistics.” *Psychological Science* 18 (5): 414–20. doi:10.1111/j.1467-9280.2007.01915.x.

Zwicky, Arnold M. 1979. “Classical Malapropisms.” *Language Sciences* 1 (2): 339–48. doi:10.1016/S0388-0001(79)80022-4.

CURRICULUM VITAE

Pedro Rui do Espírito Santo Abreu

Born	Lisbon (Portugal), 27 December 1982
Office	Instituto de Filosofia da Linguagem Av. de Berna, 26-C, Edifício I&D, Gabinete 4.02
Phone	Office +351 217908300 Mobile +351 919331613
E-mail	pedroesabreu@gmail.com
Webpage	http://www.ifl.pt/index.php?id1=2&id2=73
Position	Doctoral Candidate at the Faculty of Social and Human Sciences – New University of Lisbon; Associate Member of IFL (Institute for Philosophy of Language) – New University of Lisbon

Education

2014	Visiting Research Student in Rutgers, the state university of New Jersey, from March to May, under the supervision of Prof. Ernest Lepore
2010 -	Doctoral Studies in Philosophy, Faculty of Social and Human Sciences – New University of Lisbon
2003 - 2009	Licentiate/Bachelor in Philosophy, Faculty of Social and Human Sciences – New University of Lisbon
2000 - 2003	Bachelor in Design, ESAD (School of Arts and Design) Caldas da Rainha – Polytechnic Institute of Leiria

Scholarships and Grants

2017 -	Scholarship from the project “Ecological reasoning and decision making in innovation-oriented industry sectors at the periphery of Europe” (TUBITAK/0010/2014).

2015-2016 Scholarship in Philosophy of Language
I&D FCSH / IFILNOVA / FIL/00183

2011 - 2015 PhD Scholarship from the FCT (Foundation for Science and Technology) - Ministry of Science, Technology and Higher Education). SFRH / BD / 72674 / 2010

2008 - 2009 Integration into Research Grant (BII) from the FCT (Foundation for Science and Technology) - Ministry of Science, Technology and Higher Education)

Talks

2017 **Indeterminacy, Underdetermination, and Principle of Charity**
PhiLang 2017 - University of Lodz, Poland
13/May

2017 **Indeterminacy within Charitable Limits**
Values Research Seminar Meeting
5/May

2017 **The Constitutive Radical Interpretability of Thoughts and Meanings**
The Legacy of Donald Davidson, Graduate Conference - York University, Toronto
27/April

2016 **Rationality and Charity**
Values Research Seminar Meeting
27/Outubro

2016 **Rationalization and Convention**
Arglab Research Colloquium
14/Junho

2014 ***Essentially but not Exhaustively Rule-Governed***
Prague International Colloquium, Nature of Rules and Normativity – 18/Setembro

2014 ***Essentially Rule-Governed within Reasonable Limits***
Cracow Workshops in Analytical Philosophy, Normativity of Meaning, Belief and Knowledge – Polish Academy of Sciences, Jagiellonian University Institute of Philosophy and Department of Legal Theory – 6/Junho

2012 ***Available Evidence for Interpretation***
10th National Conference of the Italian Society for Analytic Philosophy (SIFA) – 15/September

2012 ***What counts as available evidence?***
1st IFILNOVA Graduate Conference in Analytic Philosophy – 1/Junho

2011 ***Davidson's Anomalous Monism***
Philo-D, IFILNOVA – 14/Fevereiro

2011 ***Davidson and Possible Interpretations***
Contemporary Philosophy Research in Progress – IFILNOVA – 28/Julho

Seminars and Reading Groups

2013-14	Co-lecturer (with Erich Rast) in an optional seminar for master students - *Problems of Meaning*. Faculty of Social and Human Sciences – New University of Lisbon
2011	*Reading Group on Propositional Attitudes* – Five talks about five classic papers on the topic; initiative belonging to the project *Knowledge and Ability* – IFL/UNL

Professional Experience

2003 -	Freelance Designer
2005-2010	Semente, Projectos para Audiovisuais - Designer

Languages

Portuguese	Native speaker
English	Excellent in speaking and reading, advanced in writing
French	Intermediate in speaking, advanced in reading, basic in writing
Spanish	Intermediate in speaking, advanced in reading, basic in writing