Universidade de São Paulo
Faculdade de Filosofia Letras e Ciências Humanas
Departamento de filosofia

Semantics and Communication: Essays on the Semantics-Pragmatics Divide

Tese de doutorado apresentada ao Departamento de Filosofia

Candidato:
Daniel A. Nagase
Orientador:
Edélcio Gonçalves de Souza

São Paulo
2021

Daniel Arvage Nagase

Semantics and Communication: Essays on the Semantics-Pragmatics Divide

Tese apresentada ao Programa de Pós-Graduação em Filosofia da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, para obtenção do título de Doutor em Filosofia sob a orientação do Prof. Dr. Edélcio Gonçalves de Souza

São Paulo
2021

Catalogação na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

N147s

Nagase, Daniel
Semantics and Communication: Essays on the Semantics-Pragmatics Divide / Daniel Nagase; orientador Edélcio Souza - São Paulo, 2021.
136 f.

Tese (Doutorado)- Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo. Departamento de Filosofia. Área de concentração: Filosofia.

1. Filosofia. 2. Filosofia Analítica. 3. Verdade. 4. Semântica. 5. Pragmática. I. Souza, Edélcio, orient. II. Título.

Nagase, D. A. **Semantics and Communication: Essays on the Semantics-Pragmatics Divide**. 2021. Tese. Faculdade de Filosofia, Letras e Ciências Humanas. Departamento de Filosofia, Universidade de São Paulo, São Paulo, 2021.

Aprovado em:

Banca Examinadora

Prof. Dr. Ernesto Perini Frizzera da Mota Santos Instituição: Universade de Brasília
Julgamento: _____ Assinatura: _____
Prof. Dr. João Vergílio Gallerani Cuter Instituição Universidade de Campinas
Julgamento: _____ Assinatura: _____
Prof. Dr. Ludovic Soutif Insituição: Universidade de São Paulo
Julgamento: _____ Assinatura: _____

To my grandpa and grandma:
Miguel and Naides

Acknowledgments

This thesis was written during difficult times. During what can only be described as an episode of collective irrationality, our country elected a corrupt and homicidal maniac as president. The results were predictably bad, and turned worse with the pandemic that ravaged the world in the last years. Instead of pursuing a minimally rational path, our president chose to actively pursue the death of Brazilian citizens, with the result that over six hundred thousand people have died as I write this. Paraphrasing Buffalo Nichols, maybe I am just angry and my words may be cliche, but it is hard to write a thesis while folks get murdered every day. That our times are so bleak make it all the more imperative to acknowledge those who stood with us during these hardships.

I will start, then, as I should, by saying a big, heartfelt thank you to my supervisor, Edélcio Gonçalves de Souza. The pandemic brought out the worst in some people, and Edélcio was there to support me when I had to put up with some academic shenanigans. He unflinchingly defended me and accepted to supervise my thesis in rather unfavorable circumstances, to which I am extremely grateful. Rousseau once complained, in our decadent society, to know one's friend thoroughly, one would need to wait for an emergency, that is, he thought, until it were too late. Well, in Edélcio's case, it was definitely not too late, and I suppose even Rousseau could admire the strength of character he showed then.

Speaking of friends, I would really like to take this opportunity to thank my great friend, Eduardo Marchesan. On a personal level, he was a constant presence during the pandemic, thus helping to keep me sane. Academically, he not only introduced me to the two greatest influences on this thesis, namely Chomsky and Travis, but also showed me *how* to read them, and, especially, how to read them *together*. It is not an exaggeration to say that the main ideas behind this thesis grew directly from conversations with him, in many cases out of points made *by* him. If that were not enough, he also co-organized with me two colloquia that were fundamental for this thesis, one on Kaplan in 2019 and another on Quine in 2021. Obviously, he is not to be held responsible for my particular interpretation of his ideas, nor for the use that I am here putting them.

Soon after I met Eduardo, we were joined in our discussion group by Carlos Mario Márquez Sosa. Given Carlos's engagement with the works of Gareth Evans and John Mc-

Dowell, this led to many heated debates about the cogency of Travis's criticism of them, particularly with regards to the Generality Constraint and the content of perception, and the tenacity with which Carlos defended them certainly made me better appreciate their position. I definitely tip my hat off for Carlos for holding his own in those debates, especially since he was outnumbered!

But that situation did not last long, since we were eventually joined by João Lucas Pimenta da Silva Pinto, who in many cases sided with Carlos in those debates. That was no small addition, since João's critical acumen and attention for detail made him a valuable ally. I am very happy for having him in our group, as his contributions definitely shaped my thinking on these issues. João and Carlos also read the first two chapters of this thesis, and even offered to read the rest, an offer I only did not take up because I was not fast enough to produce a workable draft for them before my deadline arrived. That was a shame, since their comments on the first two chapters were very helpful and, I hope, led to improvements on them. Moreover, they co-organized with me and Eduardo a series of lectures at USP on Perceptual Content and Singular Thought which was the occasion for some terrific debates. It is often said that writing is a solitary endeavor, and there are of course moments when one must sit alone in front of one's computer screen, but I personally had the good luck of having these wonderful companions that made writing seem to be a much more collaborative work.

In this same vein, I must also thank Daniel Peluso Guilhermino, who participated in a discussion group about Selllars with me, Eduardo, and Carlos in 2020. Although Sellars does not figure in this thesis, many of the topics we discussed there served as the background against which my thinking developed. His attention to the detail of Sellars's texts, as well as his insistence on discussing the broader implications of some of Sellars's moves (especially his appeal to sensations), made all that very rewarding. I hope we may continue this fruitful collaboration about Sellars in the future.

I have to mention also all the great folks from the logic group: Fernanda, Euclides, Levi, Marco, Taimara, Pedro Falcão, Thiago, and Luiza. Logic is famous for being a dry and arid subject, but you folks make it all very lively and fun! I hope I have contributed as much to your *Bildung* as you have contributed to mine. Those years of studying logic (and math, more generally) together made it all worthwhile; indeed, when my wife once asked me what I would do if we suddenly became rich, my immediate answer was "keep going to the logic discussion group"! I must also single out here Fernanda and Pedro Nagem for their companionship during the pandemic. Thanks, folks!

One of my oldest friends from the undergrad years is Fábio Franco. Even if, academically, we drifted apart, we nonetheless kept a strong friendship, and some of our conversations were very influential for this thesis—specifically, when discussing a version of the

second chapter, Fábio's observations regarding Lacan and the code model of language made me realize the importance of this model for discussing Lewis, thus leading to a much improved version of the chapter. That we can keep our conversation going in spite of our very different approaches is a testament to Fábio's intellectual breadth and part of our commitment against the increasing tribalization of our discipline.

Another close friend from undergrad is Dario Negreiros, who, through our numerous conversations about the nature of ideology, about the nefarious role of slavery in the history of our country, and about what it means for a country to undergo a *formation* process, also contributed immensely for my *own* formation process and again for my own fight against hyper-specialization (though, by looking at the titles of my chapters, not to mention the language in which I am writing this thesis, he no doubt will think that I have failed in my fight against colonialism). It's funny that, although Fábio, Dario, and I work in very different areas and have very different approaches to philosophy, we also share some fundamental commitments about how to pursue a philosophical project: detailed dialogue with more empirical disciplines (sociology, history, psychology, linguistics), inter-disciplinary work, collaborative efforts, a certain humbleness in *listening to others*. The older we get, and the farther apart we come in terms of the space of sub-disciplines in philosophy, the more I come to appreciate these more fundamental similarities. I am very fortunate to have friends and intellectual partners like these.

Rodrigo Freire has also supported and encouraged me through the whole process, also helping me keep in mind some of the implications of this type of discussion for the philosophy of logic and mathematics. It is a shame that the pandemics dashed some of our plans for a closer collaboration on this front, but I do hope we may be able to implement that project in the near future.

I am also extremely grateful for the members of my qualifying examination committee, David Zapero and Ludovic Soutif, for their very detailed comments and criticism of a distant version of Chapter 2. I am sure they will still find much to fault with the final text, but I am also sure they would find *much more* if they had not intervened before.

I mentioned a couple of paragraphs back that one of my fundamental commitments was to the importance of a dialogue with the empirical disciplines in order to do philosophy. This commitment has a name, to wit, *natural philosophy*, and few people have done more to promote it recently than Penelope Maddy. Prof. Maddy is one of the most generous philosophers I have met, and an email exchange with her in late 2020 was crucial for my increasing appreciation of the relevance of the literature on language acquisition for my thesis. I hope to one day join ranks with her as a *second philosopher*, as she calls those who heed the naturalistic call.

As I said, I organized with Eduardo a colloquium about Kaplan in 2019 and another

about Quine in 2021. I am very grateful for all the participants in those colloquia for the very rich discussions we had there. For the Kaplan colloquium, thanks to Marco Ruffino, David Zapero, Ernesto Perini, Pedro Santos, Thainá Dermatini, Carlos Márquez, Luiz Arthur Pagani, Rafael Albiero, and Ludovic Soutif, as well as the audience at CLE-UNICAMP, for the wonderful time I had there. For the Quine colloquium, thanks to Sandra Laugier, Gary Kemp, Gary Ebbs, Gila Sher, André Carus, Thomas Uebel, Sander Verhaegh, Paul Gregory, and the online audience for the very rich debates we had. In particular, the first colloquium was instrumental for some of my thoughts on Kaplan, which were seeds from which this thesis sprouted, and the second colloquium was crucial for the first chapter on Quine. I must also mention in this regard Prof. Dagfinn Føllesdal who, though he could not attend the Quine colloquium due to unfortunate personal circumstances, was nevertheless very encouraging about the event.

Similarly, I would also like to thank all the participants in our series of lectures about Perceptual Content and Singular Thought for their outstanding contributions: Santiago Echeverri, Mariela Aguilera, David Zapero, Susanna Siegel, Gabriel Greenberg, Bill Brewer, and Charles Travis. Together with the online audiences, we all had a great time discussing their contributions.

On a more personal level, I would like to thank Márcio Mauá Chaves Ferreira and Lucas Müssnich for being great friends, and even for offering me some reassurance about the conclusion. Thanks also to Daniel Monteiro for introducing me to the essayist tradition of Johnson, Lamb, and Hazlitt—though I am sure he will be disappointed that I have not followed them more closely stylistically. Alas!

Last but not least, I must mention my family for their unfaltering support for my intellectual pursuits, from when I was a child to my more mature years. My uncles, Miguel and Paulo, my aunts, Liliana and Simone, and my cousins, Camila (and her husband, Marcelo!), Caio, Tiago, Giovanna, Mateus, Lucas, Miguel, Kathelyn, and Isaque, are always a joy to be with. It is incredibly cliche to complain about one's in laws, but that is definitely not the case, here. Ana, Victor, Ary, Marina, I am honored that you have accepted me as part of the family. Many thanks also to Eduardo, for being a great uncle when I most needed, and to his brother, Beto, and to Daisy, for their enthusiasm for my research!

My cats, George, Harry, and Zig were faithful companions through many sleepless nights. They were incredibly solicitous, warming me when cold, jumping on me when I was falling asleep on my computer, and even trying to type some of the chapters for me on occasion. Life would not be the same without those little furry balls.

To my mother, what can I say? We have been through a lot since I was a kid, and she showed me the strength to continue on in spite of all. She taught me to read, and would not give up on me when I had given up on me. She insisted that I concluded my studies when

I did not want to go on, and now here I am, completing a Ph. D. I hope I made you proud, mom!

I started this acknowledgment mentioning my anger. There was a time when that sentiment dominated my life, and I even defined myself by the word "revolt". If I have changed, that is thanks to one person, namely my wife, Olívia Fortes, who made me into a better person and who introduced the words "happiness", "joy", and "peace" into my vocabulary. She is the one person I hold co-responsible for this thesis, as there is no way I could have done it without her. It is thus fitting that I am finishing this on our ten year anniversary.

*

Oh, and thanks to CAPES for the funding that made this research possible!

We feel that even when *all possible* scientific questions have been answered, the problems of life remain completely untouched. Of course there are then no questions left, and this itself is the answer.
—Ludwig Wittgenstein

Every time one man says to another, "Tell us plainly what you mean?" he is assuming the infallibility of language: that is to say, he is assuming that there is a perfect scheme of verbal expression for all the internal moods and meanings of men. (...) [Man] knows that there are in the soul tints more bewildering, more numberless, and more nameless than the colours of an autumn forest; he knows that there are abroad in the world and doing strange and terrible service in it crimes that have never been condemned and virtues that have never been christened. Yet he seriously believes that these things can every one of them, in all their tones and semi-tones, in all their blends and unions, be accurately represented by an arbitrary system of grunts and squeals. He believes that an ordinary civilized stockbroker can really produce out of his own inside noises which denote all the mysteries of memory and all the agonies of desire.
—G. K. Chesterston

Resumo

Nagase, Daniel Arvage. *Semantics and Communication: Essays on the Semantics-Pragmatics Divide*. 2021. Tese (Doutorado) – Faculdade de Filosofia, Letras e Ciências Humanas. Departamento de Filosofia, Universidade de São Paulo, São Paulo, 2021.

A presente tese é um estudo aprofundado da assim chamada tradição interpretativa, consituída principalmente por Quine, Davidson e Lewis. Nosso objetivo é interrogar essa tradição a partir da divisão semântica-pragmática, realçando certas tensões no interior da obra desses autores. O primeiro capítulo, assim, analisa a obra de Quine, procurando mostrar que uma certa leitura, a qual vê em Quine um defensor da publicidade do significado, se fundamenta em uma confusão dos níveis de explicação propostos pelo mesmo. O segundo capítulo analisa a obra de Lewis, procurando mostrar que seu modelo explicativo, o qual vê as sentenças de uma linguagem como códigos para informações sobre o mundo e atividade comunicativa como simplesmente um procedimento de codificação e decodificação, é por demais empobrecido para dar conta da complexidade do uso da linguagem. De fato, o próprio Lewis fornece as ferramentas para se pensar um modelo mais completo a partir de sua noção da atividade linguística como uma ferramenta para a coordenação de ações racionais. Por fim, o último capítulo se dedica a obra de Davidson, explorando como, em um primeiro momento de sua obra, ele procurava fundamentar o significado linguístico em convenções linguísticas, para, em segundo momento de sua obra, rejeitar essa explicação em prol de uma concepção mais complexa da situação comunicativa. Nossa conclusão é dupla. Por um lado, a semântica, como o estudo das características estáveis de enunciados linguísticos, deve ser reconfigurada, restringido-se em grande parte à gramática. Por outro lado, a pragmática, como o estudo da interação do contexto com os enunciados linguísticos deve ser ampliada para abarcar também as condições de verdade das sentenças.

Palavras-chave: Quine – Davidson – Lewis – Semântica – Pragmática – Comunicação

Abstract

Nagase, Daniel Arvage. *Semantics and Communication: Essays on the Semantics-Pragmatics Divide*. 2021. Thesis (Doctorate Degree) – Faculdade de Filosofia, Letras e Ciências Humanas. Departamento de Filosofia, Universidade de São Paulo, São Paulo, 2021.

The present thesis is an in-depth study of the so-called interpretationist tradition, which is primarily constituted by Quine, Davidson, and Lewis. My objective is to study this tradition from the point of view of the semantics-pragmatics divide, highlighting certain tensions in the work of these authors. The first chapter deals with Quine's work, arguing against a certain reading of this work, which sees in Quine a proponent of the thesis of the publicity of meanings. The basic thrust of this chapter is that this reading rests on a misunderstanding of the levels of explanations proposed by him and the location of public meanings relative to these levels. The second chapter analyzes the work of David Lewis, arguing that his explanatory model, which sees sentences as codes for pieces of information about the world and sees the communicative activity as simply an encoding and decoding procedure, is too impoverished to deal with the complexities of language use. Indeed, Lewis himself provides us with the tools to build a more comprehensive model of language use with his notion of linguistic activities as tools for the coordination of rational action. Finally, the last chapter analyzes Davidson's work, showing that if, at first, he based linguistic meaning on linguistic conventions, in a second moment, he rejected this explanation in favor of a more complex conception of the communicative situation. My conclusion is twofold. On the one hand, semantics, as the study of the stable characteristics of linguistic utterances, must be reconfigured, being now restricted mostly to grammar. On the other hand, pragmatics, as the study of the interactions between utterances and contexts, must be broadened so as to also include the study of the truth conditions of sentences.

Key words: Quine – Davidson – Lewis – Semantics – Pragmatics – Communication

Contents

Introduction

This thesis is, in a certain sense, the chronicle of a mistake. The mistake is to expect too much from semantics. A special case of this is to expect semantics to deliver truth conditions for sentences in context, a mistake we see with particular clarity in the works of Davidson and Lewis. The root of this mistake can, perhaps, be traced to Quine, in particular to a certain misreading of him to the effect that meanings are public entities that do explanatory work when considering how communication can be successful.

Appropriately, then, this chronicle begins with Quine. The first chapter of this thesis examines the idea that Quine defends a version of the Publicity Thesis, i.e., that meanings are public entities. In particular, it examines two supposed Quinean arguments for this thesis, namely the Learnability Argument—the conditions for language learning presuppose that there are meanings as public entities—and the Communication Argument—the conditions for successful communication presuppose that there are meanings as public entities. Unsurprisingly, given that this is a chronicle of a mistake, it finds both these arguments wanting. The source of the mistake is in both cases the same, to wit, to suppose that the coordination of behavior that founds the supposed publicity of meaning is actually explanatory. Against this, I point out how Quine himself always had it clear that this coordination is *what must be explained*, and that his explanation for that does not invoke any mysterious meanings as public "entities" (as Quine would undoubtedly insist, meanings are not *entities* at all, public or otherwise).

One thing that emerges out of this analysis is that part of what does the explaining is our neurophysiological endowment. Now, as is well known, Quine is often distrustful of Chomskyan speculation regarding the linguistic component of this endowment, preferring, where possible, to stick with general cognitive mechanisms. Still, even he is forced to admit (Quine 1968/1975) that there must be a specific mechanism responsible for our linguistic competence, one that is not reducible merely to general mechanisms that we share with our primate cousins (otherwise, it would be hard to explain why *they* do not develop the same linguistic competence as us, even under strenuous training). And even if Quine himself ultimately proved to be skeptical of the extent of our innate linguistic competence, the Quinean naturalist should not be, since there is considerable empirical evidence supporting

a rather rich language faculty.[1] This is, perhaps, the first general lesson I would like a reader to take out of this chronicle: ignore our innate linguistic endowment at your peril!

This naturally leads to the second chapter, where I examine in detail how Lewis ignored this innate linguistic endowment at *his* peril. I argue there that, as much as he wanted to relegate our human linguistic competence to a marginal place in an account of the nature of language and meaning, he is nevertheless forced to concede a significant role to it in his account. Part of the problem has to do with the mistake I identified in Quine, namely that Lewis confuses the levels at which linguistic explanations take place. In particular, he believes that our common sense notions about meanings can be turned into a systematic account of the contributions of each lexical item to the truth conditions of the sentences in which it appears. This presupposes that these truth conditions are *stable*, and that therefore they cannot vary depending on the goals and purposes to which the sentences are put. As I argue in this chapter, this presupposition is untenable.

Ironically, Lewis himself gives us the tool to see that this cannot be so, since he insists that communication is a rational collaborative activity—a tool for solving coordinate problems, as he terms it. Taking a cue from Herbert Clark (1996) persuasively shows, I argue that this means that we must see how the assignment of truth conditions to utterances *is itself a collaborative activity*. The net result is that, first, if common sense deals with truth conditions, then these truth conditions cannot be systematized in a general semantics, as Lewis proposed. In turn, freed from the straight-jacket of giving a systematic account of meaning that is consonant with common sense, we are able to give grammar its proper due in determining the truth conditions of utterances, as well as making room for the social collaborative activity—communication—as part of the story of how truth conditions are assigned to utterances. That is the second general lesson the reader should take from this chronicle: to ignore the semantics-pragmatics divide is to fail to do justice to both, while restoring allows us to see them in their proper light.

And so we come into the third and last chapter, in which these issues come into sharper focus. Davidson is perhaps most famous for his proposal that a proper semantic theory for a language should take the form of an axiomatic theory of truth for the language in question. During his career, however, he struggled to give a proper justification for this idea. At first, he tried to justify it in terms of linguistic conventions: an axiomatic theory of truth would capture the *conventional semantic core* of a language. Later, though, he rightly came to see that this could not be right, since interpreters often need to attribute idiosyncratic truth conditions to a speaker. Unfortunately, he did not follow this realization to its end, and still held that a semantic theory for a language should take the form of an axiomatic truth theory. The chapter explores this tension, arguing that it is only by abandoning the ideal of an

[1] See especially the phenomenon of *homesign* as analyzed in that chapter.

axiomatic truth theory that we can better appreciate Davidson's insight that communication proceeds by the participants *negotiating* the truth conditions of their utterances. Part of the reason Davidson still clung to an axiomatic truth theory is because he saw no other way of explaining the *systematic* character of our linguistic competence. But I will argue there that the systematic character is best seen as coming from *grammar*, not from truth conditions. That may be the final lesson of this thesis, to wit, that the stable and systematic component of language comes from grammar, whereas the determination of truth conditions is a social, cooperative process open to negotiation.

The above summary may give the reader the impression that each chapter deals with a specific theme, with the thesis moving unrelentlessly forward towards its conclusion. That is not quite true. The thesis may be unrelenting—I will let the reader judge that!—but it is not as orderly as the above may imply. I did mean each chapter to be more or less autonomous, and this entailed that there may be some repetition and overlapping among them. So the lessons which I associated with specific chapters may be found spread out among all of them. The chapters are also not very orderly, sometimes with length digressions into aspects of the view of the author being analyzed in the chapter in question. That is why I chose to call them *essays*, though, truth be told, they do not have much in common with the writings of Hume or Johnson, much less of Hazlitt and Lamb. It does sound better to call them essays than calling them "chaotic mess", so I stuck with that.

Chapter 1

A Quinean Argument for Meaning Internalism

1.1 Introduction

Quine is generally taken to be a paradigmatic defender of the idea that meanings are inherently public. Indeed, *Word and Object* opens with the declaration that: "Language is a social art. In acquiring it we have to depend entirely on intersubjectively available cues as to what to say and when. Hence there is no justification for collating linguistic meanings, unless in terms of men's dispositions to respond overtly to socially observable stimulations" (Quine 1960/2013, p. xxix). Passages such as this apparently justify attributing to Quine the following thesis:

Publicity Thesis: Meanings, whatever they are, must be public entities, i.e. they must be intersubjectively available.

In spite of the impressive rhetoric, however, I will argue that a properly Quinean philosophy of language must instead conclude that meanings are best conceived as internalistic entites. In other words, I will defend that one should conclude, from Quinean premises, *meaning internalism*:

Meaning Internalism: Meanings are best conceived as private psychological states.

This may seem outrageous. Have I not just quoted Quine as saying that meanings can only be thought of "in terms of men's dispositions to respond overtly to socially observable stimulations"? If that is a requirement on meanings, surely they cannot be conceived in terms of private psychological states, which are precisely *not* overt responses to socially

observable stimulations. My strategy will be to argue that this reaction rests on a misunderstanding. True, Quine often construes meanings in terms of dispositions to behave in determinate ways in certain situations. However, this construal corresponds to *the phenomenon to be explained*, not to what does the explaining. More specifically, Quine holds that these dispositions must eventually give way to neurophysiological explanations. But neurophysiological explanations are not public; indeed, under the reasonable assumption of token physicalism, they are private psychological states.[1]

So, on the one hand, we have the superficial level of *explananda*, and here meanings are indeed public. But on the deeper level of the *explanans*, we have certain private psychological states that underlie the behavioral dispositions in question. My proposal is that there are good reasons for calling such states "meanings", irrespective of whether they share some or all of the characteristics typically associated with the phenomenon located at the superficial level. In Quinean jargon (Quine 1960/2013, p. 238), the reconstruction of meanings as private psychological states, as defended by Meaning Internalism, is an *explication* of the notion of meaning, that is, a recruitment of this term for scientific purposes. It therefore does not have to capture all the dimensions of the superficial notion—on the contrary, that would defeat its purpose as a scientific *replacement* for the common notion.[2]

This essay is therefore structured as follows. First, I will discuss two types of arguments commonly associated with the Publicity Thesis, in order to show that they founder even if considered through a Quinean lens. This will lead into the distinction between the level of what must be explained and of what does the explaining, as associated with Quine's empiricism, which in turn will lead us, in the next section, into a discussion of Quine's views on scientific explanation. Finally, I will show how Quine himself appealed to private psychological states in his own explanation of the phenomenon behind the Publicity Thesis, and how this can be used to motivate a reconstruction of the notion of meaning along

[1] My position is thus very similar to the one defended by Kemp (2017), an article that strongly influenced my thinking. My main difference with Kemp is that I believe there are strong empirical reasons for preferring the Chomskyan treatment of syntax to Quine's, and that, in any case, there is nothing in Quine that precludes the Chomskyan treatment (though, if I am reading him right, nothing Kemp says indicates that he believes the Quinean should not be a Chomskyan; rather, he is simply saying that, as a matter of historical fact, Quine preferred other types of explanations). Drobňak (2018) also defends that Quine is closer to internalism, but, like Kemp, does not emphasize the empirical evidence as I do. Moreover, Drobňak thinks that Quine's stance privileges the *idiolect*, and, though I will not argue the point here, this seems misguided—neither the idiolect nor languages should be considered as fundamental, but, instead, what is in focus is the individual's linguistic competence. Cf. Collins (2010, p. 47) and Begby (2016, p. 46n39) for this latter point.

[2] And so, I believe Quine would part company with Lewis (1974/1983), for whom the term meaning is a theoretical term—in the sense of Lewis (1970/1983b)—which appears in a rough formalization of common sense, and thus for whom philosophical analysis of this term cannot consist of anything more than an elucidation of the "platitudes" which common sense associates with it.

internalist lines.

1.2 Publicity

There are two types of arguments for the Publicity Thesis in Quine's works. One is a *learnability argument*, that is, an argument which takes as its starting point the conditions under which we learn a language, and proceeds to infer from these learnability conditions the Publicity thesis. The other is a *communication argument*, which takes as its starting point the conditions under which successful communication takes place and infers from such communication conditions the Publicity Thesis. As much as those arguments find support in Quine, in the end, I will conclude that both arguments are unsound *even by Quinean lights*. Perhaps this just points to an inconsistency in Quine,[3] but I believe that it is possible to give a more coherent reading of such passages once one puts them in the larger context of Quine's naturalism.[4]

As I see it, the deeper point unearthed by this discussion is twofold. First, it is clear that even a Quinean scientific account of language cannot do without private psychological states, suitably interpreted. This is why the two arguments discussed above founder even by Quinean lights. Second, and more importantly, these arguments confuse the level at which these Quinean passages are located. That is, they consider talk of meaning as being *explanatory*, and thus miss the point of Quine's position: meanings are not public in the sense of being an explanatory posit in a scientific account which happens to be public, rather, meanings are public in the sense that they are reducible to behavioral dispositions which, in their turn, are merely symptoms of underlying psychological states which are the real explanatory posits. In other words, meanings, in the sense in which they are public, are what must be explained, not what does the explaining.[5]

[3] This is Chomsky's (1975, pp. 198ff) position.

[4] Jankovic and Ray (2017) find these two strands in Davidson, attributing them only in inchoate (or "aphoristic") form to Quine. In particular, they find it difficult to pry these strands apart from his "commitments to naturalism, behaviorism and a verificationist theory of meaning" (Jankovic and Ray 2017, p. 100). Given what I just said, perhaps the difficulty in question should be seen as the result of Quine *not* subscribing to those arguments, so that, if Davidson was influenced by these passages in Quine, it is quite possible he simply misread them. This would not be too surprising, given their differences—cf. Kemp (2012) for a Quinean assessment of Davidson's project.

[5] Indeed, I believe this illustrates a larger trend in Quine, namely to consider public facts as *explananda*, not as *explanans*. That is why he never fully embraced Davidson's triangulation argument, since this latter argument *presupposes* what must be explained, a point that was unfortunately missed by, e.g., Føllesdal (1995, p. 59).

1.2.1 The Learnability Argument

Let us first analyze the *learnability argument*, as some have found in this famous passage from *The Pursuit of Truth*:[6]

> In psychology one may or may not be a behaviorist, but in linguistics one has no choice. Each of us learns his language by observing other people's verbal behavior and having his own faltering verbal behavior observed and reinforced or corrected by others. We depend strictly on overt behavior in observable situations. As long as our command of our language fits all external checkpoints, where our utterance or our reaction to someone's utterance can be appraised in the light of some shared situation, so long all is well. Our mental life between checkpoints is indifferent to our rating as a master of the language. There is nothing in linguistic meaning beyond what is to be gleaned from overt behavior in observable circumstances. (Quine 1992, pp. 37–8)

Dorit Bar-On (1992b, p. 241) reconstructs the argument roughly like this:

(P1) We come to know the meaning of our words by observing the behavior of other people in observable circumstances.

(C1) Therefore, in coming to know the meaning of our words, "we depend strictly on overt behavior in observable circumstances".

(C2) Therefore, given this strict dependence, whatever cannot be "gleaned from overt behavior in observable circumstances" cannot also play a role in our coming to know the meaning of our words.

(P2) But whatever plays no role in our coming to know the meaning of our words cannot be part of this meaning.

(C3) Therefore, whatever cannot be "gleaned from overt behavior in observable circumstances" cannot be part of the meaning of our words.

It is clear, then, that this argument rests on two premises, namely (P1) and (P2) above. Consider first (P2). Suppose that something is part of the meaning of a given expression. Clearly, I am not born associating the expression to this meaning, so I must have learned it somehow. But then, supposing that in learning the meaning of an expression, I thereby learn that whatever is part of the meaning of an expression *is* part of this meaning, I must have learned that this something is part of the meaning of the expression, whence it played

[6] This passage also appears *verbatim* in "Indeterminacy of Translation Again" (Quine 1987, p. 5).

a role in my learning the meaning of the expression. Thus, by contraposition, if something does not play a role in my coming to know the meaning of an expression, it cannot be part of this meaning.

Notice that this argument depends on the notion that, in knowing the meaning of an expression, I know the parts of this meaning. This could be justified if knowledge of the meaning of expressions were to be *conscious* knowledge. Now, it is well known that theorists such as Chomsky *deny* that there is any kind of knowledge involved here.[7] What about Quine? Does Quine subscribe to this principle? In his correspondence with Quine, Nelson Goodman suggests that a person could come to know the meaning of a word, in the sense of understanding it, without being able to fully articulate it. Quine then replies that he thinks it is a condition on a person's knowing the meaning of two words that the person be able to reflectively establish whether two words have the same meaning.[8] This does seem to point in the direction of Quine subscribing to something like the above principle. Nevertheless, I will argue in Section 2.3 that Quine is best read as abandoning this idea. But for now, let us assume that it is plausible, and thus that (P2) does go through.

What about (P1)? This claim may seem like a platitude: of course people learn a language by observing other people's behavior; how else could they acquire a language? One idea is that someone could perhaps *create* a language even though this person has never had any contact with other beings. What is known about linguistic acquisition seems to rule out this possibility, though: the literature is unanimous about the need for *some* social interaction in order for the child to acquire a language. Moreover, not just any social interaction is necessary: this interaction must also involve *some* communication and be supportive—as Goldin-Meadow aptly puts it, the child must interact with humans "and those humans, at a minimum, must be humane" (Goldin-Meadow 2003, p. 48).[9]

But if (P1) is read as merely pointing for the necessity of *some* (humane) social interaction for the acquisition of language, then the argument does not go through. In order for the argument to go through, (P1) must require not only that social and linguistic interactions take place, but also that such interactions *completely and exhaustively* structure the linguistic output of the child. In other words, the observation of other people's behavior mentioned in (P1) must be the *sole* source of our knowledge of the meaning of our words. Readers of

[7] Cf. George (1987, pp. 161) and Collins (2004) for discussion.

[8] Goodman's letter, and Quine's reply, are reproduced in Verhaegh (2018, pp. 177-180).

[9] For textbook presentations of this necessity, including a discussion of the infamous case of Genie, cf. Hoff (2013, chap. 2), Rowland (2014, pp. 16), and Saxton (2017, chap. 3). More abstractly, Bar-On (1992a) considers the possibility of a "Super-Crusoe", a being that is able to invent a language without any contact with other beings, and argues against the plausibility of such a being by exploring what she sees as the centrality of the communicative function of language. This seem reasonable enough—cf., e.g., Tomasello (2008)—, but do note that even here there are doubts—cf. Hauser et al. (2002) and Reboul (2017).

Quine could find some support for this idea in his speculations about the learning mechanisms behind language acquisition, namely reinforcement and imitation (Quine 1975, p. 84).[10] If the child only acquires a language by being either reinforced into a pre-existing linguistic pattern or by imitating such a pattern, then the only source for her knowledge of the meaning of her words is indeed other people's behavior. So here, perhaps, one can see how a broadly behaviorist theory of language learning could support this premise.[11] Are these, however, the only learning mechanisms used by children to acquire a language? This is an empirical question, and there are good empirical reasons for holding that this is *not* the case.

In their remarkable studies of deaf children of hearing parents, Feldman et al. (1978) and Goldin-Meadow (2003) show how these children effectively create their own languages (called *homesigns*) in order to communicate with others.[12] Since the children could not hear their parents speaking, and the parents, for many reasons, did not communicate with them by using conventional sign languages, these children could not have invented their languages *solely* by observing other people's behavior. One could object that the children came up with their homesign by observing their parents' *gestures*, but Goldin-Meadow (2003, chap. 14) shows that there are so many differences between these gestures and the children's gestures that this is simply not plausible. Indeed, whereas the children's gestures formed a *linguistic system*, with a stable lexicon, rudimentary syntax, and recursive devices, the parent's gestures never attained this degree of systematicity. Strikingly, by comparing the emergence of homesign in Chinese and American homes, Goldin-Meadow (2003, chap. 15) also shows that the children's gestures resemble more *other homesign* systems than the gestures of their parents (by, for instance, showing a preference for ergative, instead of accusative, patterns of word order).[13] Unsurprisingly, Goldin-Meadow takes this conver-

[10]It is a bit surprising to see Quine appealing to those mechanisms, and especially to the importance of random babbling, in his theory of language acquisition. Already Lenneberg (1964, pp. 598, 601–603) had emphasized the insufficiency of these mechanisms, e.g., for the language acquisition of deaf or mute children.

[11]If this is indeed Quine's position, then one must disagree with Hylton (2007, pp. 103) when he says that Quine is here appealing merely to the "undeniable fact" that "language is learnt by infants who receive the information about the world only through their sensory s[t]imulations", so that Quine is not here "committed to an outmoded approach to psychology". That is, if (P1) is supported by a further hypothesis about the learning mechanisms of the child, then it is not just a platitudinous commitment to a broad empiricism. Furthermore, it is not even clear that this empiricism is indeed platitudinous or an "undeniable fact", since it is conceivable a person may also have received information about the world *from her innate cognitive endowment*, something that I am not sure Quine wishes to deny.

[12]The philosophical literature on homesigns is rather scarce, but do see Jackendoff (2003, pp. 99ff) and Begby (2017) for comments that point in the same direction as the one pursued here.

[13]Incidentally, Goldin-Meadow (2003, Part II) also goes to great lengths to show that homesigns *are* languages, i.e. they have a stable lexicon and syntax, and are not just an unstructured system of conventional signs. Given the evidence, then, I find that the claim by certain theoreticians (e.g. Hoff (2006, p. 55), Tomasello

gence as evidence for the hypothesis that children have innate (conceptual and linguistic) competences which they bring to bear in the creation of their homesigns.

The phenomenon of the homesign shows then that the observation of other people's behavior can be at most an *enabling condition* for the child to come to know the meaning of her words. That is, in the case of the child developing her own homesign, observing the behavior of other people is important in the sense that a *nurturing social environment* is important for the child's development, but *not* in the sense that the child is learning the meaning of the words *from* the observed behavior. In fact, these children are spontaneously *endowing* certain forms with meaning. Consider, for example, the following description of the way these children recruited a conventional gesture to their vocabulary:

> As another example, all of the deaf children use the "give" gesture—a flat hand, palm-up, extended to request the transfer of an object (...). When hearing children use this gesture (and most do), they use it to mean "put something in my hand." But the deaf children use this gesture to request the transfer of objects, not only to themselves, but also to other people and locations. Moreover, the gesture is used for objects that will fit in the hand but also for those that won't. The gesture thus has an extended meaning for the deaf children, one that begins to eat away at the transparency between gesture form and meaning. (Goldin-Meadow 2003, p. 76)

The last sentence of this quotation points to an important phenomenon, namely that, in being recruited into their lexicon, the children's gestures cease to be iconic and become part of a linguistic system.[14] As a consequence, the gesture is now interpreted inside a predicate frame structure, i.e. an underlying argument structure with thematic roles. In the above case, for instance, the gesture assumes the thematic structure of, say, "ACTOR transfers OBJECT to RECIPIENT", which is present even in the cases where the child does not make all the arguments explicit (by deleting some of them).[15] Notice that, for this to be the case, these thematic roles must be previously available to the child. Obviously, they are not present at birth, but they are not learned either; rather, it is part of the child's *maturational development* to eventually acquire these structures.[16]

(2008, pp. 257ff), and Morgan (2005, pp. 926ff)) that these are either "deficient" or, indeed, not really languages at all, as bizarre *ad hoc* attempts to deflate the phenomenon.

[14]That is, they become *iconically motivated symbols*, in the helpful terminology of Greenberg (n.d.). This means that the *semantic* connection between symbol and meaning is arbitrary, even if historically the choice of the symbol was iconically motivated.

[15]Cf. Goldin-Meadow (2003, chap. 10) for an extensive argument for the presence of thematic roles as predicate frames in the deaf children's homesigns.

[16]This perspective on language acquisition was famously defended by Lenneberg (1964, 1967) and even-

Now, one may think that these children's situations are far from typical, but, in fact, this process of *endowing* certain verb forms with meaning by using this predicate frame structure is a widespread strategy in learning a language, a process known as *syntactic bootstrapping*.[17] Consider the example of Kelli, a congenitally blind child studied by Landau and Gleitman (1985). Being blind, one would expect Kelli to have difficulty learning sight related verbs such as "see" and "look". Yet not only did Kelli use them on her own way, but she also displayed mastery of subtle distinctions, such as of the fact that "see" is stative and "look" is active. How did she get such knowledge? Landau and Gleitman conjectured that she managed to do so at least in part by analyzing the predicate frames on which these verbs appeared, and therefore homing in on the meaning of the verb. Again, in order to do so, Kelli must have already had these predicate frames available, so a reasonable hypothesis is that they were acquired as part of her natural linguistic development.[18]

The empirical data thus points towards the falsity of (P1). Observation of other people's behavior does not exhaust the source of the child's acquisition of the meaning of the words. Some syntactic knowledge, perhaps in the form of an innate grammar, is also an important factor for language acquisition, and this factor drives the way children assign meanings to the words in their lexicon. But are such learning mechanisms allowed by Quine's strictures? I believe that something similar was proposed by Quine in his account of how children learn new words by "analogy" with old ones:

> It is evident how new sentences may be built from old materials and volunteered on appropriate occasions simply by virtue of the analogies. Having been directly conditioned to the appropriate use of 'Foot' (or 'This is my foot') as a sentence, and 'Hand' likewise, and 'My foot hurts' as a whole, the child might conceivably utter 'My hand hurts' on an appropriate occasion, though unaided by previous experience with that actual sentence. (Quine 1960/2013, p. 8)[19]

Quine is not very explicit on this point, but the idea seems to be that the child notices

tually adopted by Chomsky. Cf. Collins (2007, p. 645n28) and the references therein. For current research focusing specifically on the acquisition of argument structure, cf. Borer (2004).

[17]This process was proposed by Lila Gleitman and her associates in a series of papers. Cf. in particular Gleitman (2020, Parts IV and V) and Landau and Gleitman (1985, chap. 7). Cf. also Borer (2004) for important evidence from Hebrew.

[18]Jackendoff (1983) also argues that at least some thematic roles are innate, though do note that he thinks the thematic roles are then associated with a *lexical item*, and not with the syntax. Borer (2003) has since argued that we should think of these thematic roles as emerging from predicate frames that are independently available; a lexical item, then, does not have a fixed number of arguments (indeed, in some cases, not even a fixed grammatical category, such as noun or verb), and is *coerced* into taking certain arguments by being inserted into the appropriate frame. As noted by Borer (2004), this approach is much more congenial to syntactic bootstrapping.

[19]This point is also emphasized in Quine (1973, 1995, resp. p. 42 and p. 7).

that her words have similar functions, and from that infers that they follow a similar grammatical pattern.[20] But in order to do so, the child must somehow employ the notion of a grammatical pattern in her analysis of utterances. Importantly, this grammatical pattern is not behaviorally detectable, since it is not apparent on the surface form of utterances, and so must be *imposed* by the child on the data. This is all recognized by Quine himself, who also speculates that this process may indeed be driven by "innate structures":

> Learning by ostension is learning by simple induction, and the mechanism of such learning is conditioning. But this method is notoriously incapable of carrying us far in language. This is why, on the translational side, we are soon driven to what I have called analytical hypotheses. The as yet unknown innate structures, additional to mere quality space, that are needed in language-learning, are needed specifically to get the child over this great hump that lies beyond ostension, or induction. (Quine 1968/1975, pp. 57–8)

I will come back to this point in section 2.3 in connection with the indeterminacy of translation thesis, but for now, note that this picture is very close to the process we saw above in which the children recruited a word into their linguistic system by either placing it inside a predicate frame structure, or else by using this predicate frame structure to hypothesize about the meaning of the word. Of course, Quine's description is much more abstract, not to say vague, but nevertheless it is clear that he has no objection in principle to the child's imposing a syntactical analysis on the data in order to form a coherent linguistic system. The important point, then, to extract from this discussion is this:

Quinean Nativism: A child learning a language possesses innate structures which drive her language learning processes.

It is at this point, however, that the *communication argument* intervenes. Nevermind, says Quine on this reading, that children can bring more to bear on their learning of new words than just what they pick up from others in observable circumstances. Whatever idiosyncratic strategies they use to learn the meaning of a word, such strategies are *irrelevant* to meaning, since they do not affect communication.

[20] Here, he may have been influenced by Bloomfield: "A grammatical pattern (sentence-type, construction, or substitution) is often called an *analogy*. A regular analogy permits a speaker to utter speech-forms which he has not heard; we say that he utters them *on the analogy* of similar forms he has heard." (Bloomfield 1933/1961, p. 275).

1.2.2 The Communication Argument

This argument can be apparently found in some of Quine's remarks on communication. In *Word and Object*, he explains the idea thus:

> The uniformity that unites us in communication and belief is a uniformity of resultant patterns overlying a chaotic subjective diversity of connections between words and experience. Uniformity comes where it matters socially; hence rather in point of intersubjectively conspicuous circumstances of utterance than in point of privately conspicuous ones. For an extreme illustration of the point, consider two men one of whom has normal color vision and the other of whom is color-blind as between red and green. Society has trained both men by the method noted earlier: rewarding the utterance of 'red' when the speaker is seen fixating something red, and penalizing it in the contrary case. Moreover the gross socially observable results are about alike: both men are pretty good about attributing 'red' to just the red things. But the private mechanisms by which the two men achieve these similar results are very different. The one man has learned ' red ' in association with the regulation photochemical effect. The other man has painfully learned to be stimulated to 'red' by light in various wavelengths (red and green) in company with elaborate special combinations of supplementary conditions of intensity, saturation, shape, and setting, calculated e.g. to admit fire and sunsets and to exclude grass; to admit blossoms and exclude leaves; and to admit lobsters only after boiling. (Quine 1960/2013, p. 7)

Or, as he puts it more trenchantly in *The Roots of Reference*:

> Language, we are told, serves to convey ideas. When we learn language, we learn to associate its words with the same ideas with which other speakers associate them. Now how do we know that these ideas are the same? And, so far as communication is concerned, who cares? We have all learned to apply the word 'red' to blood, tomatoes, ripe apples, and boiled lobsters. The associated idea, the associated sensation, is as may be. Language bypasses the idea and homes on the object. Than the idea there is little less useful to the study of language. (Quine 1973, p. 35)

This argument can be rendered like this:[21]

(P1) Whatever does not play a role in successful communication is not part of the meaning of the expression.

[21] A similar way of rendering the argument can be found in Fisher (2011, p. 56).

(P2) Private psychological mechanisms do not play any role in successful communication.

(C1) Therefore, private psychological mechanisms are not part of the meaning of expressions.

Let us start with (P1). The key idea seems to be this:

Communication Principle: Successful communication involves the hearer grasping the meaning of the speaker's utterance.[22]

Hence, if we can communicate without each grasping a determinate posit, then this posit is, *ipso facto*, not part of the meaning of the expressions used, since, if it was part of the meaning, it would be grasped by the interpreter. So we have (P1).

Assuming the Communication Principle, (P2) also follows. The point of (P2) is not that private psychological mechanisms play no role in the successful communication, in the sense that they are not employed either by the speaker or by the hearer; this would be plainly absurd. The point is that they may as well employ them both in the production and in the interpretation of utterances, but the point is that such private psychological mechanisms are not *what is communicated*. Since the hearer has no access to the private psychological states of the speaker, it is presumably not by grasping such private psychological states that successful communication takes place. That is, the hearer's grasp of what the speaker said may constitute a private psychological state, but the *content* of such a state is not *the speaker's* private psychological states. That is, in understanding what the other said, I may *be* in a private psychological state, but I am not thereby *grasping* a private psychological state. So, assuming that communication involves grasping the meaning of the other's utterance, the argument seems straightforward enough.

Notice that, with the Communication Principle, the Publicity Thesis is presented as an *explanation* of what constitutes successful communication. It is because meanings are public that they can be grasped by the different participants in a communicative situation, and hence that successful communication can take place. So meanings would be here theoretical posits that *explained* successful communication. Is this explanation sound, though?

Consider again the example of Kelli, the congenitally blind child studied by Landau and Gleitman (1985). There is a clear sense in which Kelli associated different things with the verbs "look" and "see" than her sighted peers: whereas her sighted peers associated these verbs roughly with "perceive by sight", she associated them roughly with "perceive by touch". This means that Kelli associated different truth conditions to utterances involving "see" and "look" than her sighted peers, since the latter would usually not count a case of

[22] This principle is clearly present in Dummett. Cf. George (1997) for a very insightful analysis of this and related ideas.

touching as a seeing, whereas she would. There is thus a strong presumption that there is nothing "public" that is grasped by both Kelli and her sighted peers in those interactions. But according to Landau and Gleitman, Kelli only noticed that she used these terms differently than her sighted peers when she was *five years old*. Until then, she had what by all accounts was successful communication with them, even in her use of "see" and "look"—indeed, it was the *breakdown* of communication which drew her attention to their differences in the use of these verbs.[23]

Presumably, we should not say that she was grasping the meaning of her peers' utterances before those incidents, and then *stopped* grasping those meanings. Conversely, it is not like she was not successfully communicating before she noticed the difference, and only began to engage in successful communication after the incidents. Rather, we should be saying that communication was possible *in spite of* there being nothing "public" that was grasped by all the participants in the conversation. And communication continued to be successful even after she realized that she employed those words differently from her sighted peers *because* communication does not depend on the participants of a conversation using the words in the same way. At this point, a proponent of the communication principle may wish to bite the bullet and simply say that Kelli was not really successfully communicating with her sighted peers. If so, however, then she would not find Quine on her side. As Quine makes clear, successful communication depends on *pragmatic* criteria:

> Success in communication is judged by smoothness of conversation, by frequent predictability of verbal and nonverbal reactions, and by coherence and plausibility of native testimony. (Quine 1992, p. 43)

I think Quine is right on this point. Communication is a joint activity, that is, a kind of activity in which two or more participants coordinate their efforts to realize some common goal that they could not achieve alone.[24] This means that success in communicating is not defined abstractedly, but rather with reference to whatever particular goals are being served by a particular communicative situation. Hence, success in communicating is not defined by the participants "grasping" some antecedently available entity either, whatever that means.

[23]This involved two separate incidents, one involving family pictures and the other a hot air balloon. The first made Kelli realized that she could tell the difference in the pictures just by feeling them, and the second made Kelli realize that others could see things that she could reach. Cf. Landau and Gleitman (1985, pp. 201ff) for details.

[24]Cf. Clark (1996, chap. 2) for more on joint activities. I am here adopting the simplifying assumption that communication is always cooperative; if the reader feels this is too much a simplification, simply replace "communication" with "collaborative communication". Even supposing there is a more encompassing account involving collaborative and adversarial communication, it seems to me desirable to have an account of collaborative communication before considering the complications introduced by adversarial communication.

Of course, both Kelli and her sighted peers associated similar things with "see" and "look", so it is not like the differences in their associations are wholly unconstrained—as the story goes, Humpty-Dumpty had a hard time communicating successfully.[25] The point is that communication involves just some kind of *coordination* among individuals, and that this coordination can take place even if participants have erroneous beliefs about the signals sent by the others.[26] This is especially so because the participants may have no *specific* beliefs about their own signals, such beliefs being acquired only as a *result* of the communicative situation at hand.

If this is so, however, then perhaps private psychological mechanisms *can* be relevant for the meaning of the expressions, even if they are not what is grasped (if anything is) when people communicate. This would require us to give up the idea that people *know the meaning* of the other person's utterances, even when they successfully communicate. Indeed, it may even require that we give up the idea that people *consciously know the meaning* of their *own* utterances, since they may not be privy to the psychological mechanisms that are coupled with their interpretation of a given expression. According to this picture, meanings, in the sense of the constant semantic features studied by linguistics, may *constrain* what is said by a given utterance, but they do not fully *determine* it, because *what is said* emerges *in the course of communication*.[27] I believe Quine himself ended up with a position which does have this consequence, but, before getting into that, I want to probe deeper into the connections between the Publicity Thesis and Quine's empiricism.

1.2.3 Quine's Empiricism

In many places, Quine suggests that acquiring a language is much like acquiring a theory, in that the child can be seen as forming *hypotheses* about the language he is learning. For example, much like the linguist must form *analytical hypotheses* about the segmentation of sentences into words, and from there proceed to devise both a grammar and a lexicon for the language she is studying, so, too, must the child form conjectures about the structure of her language and about the meaning of its words. If, however, the child is forming such a theory, then it is reasonable to ask: what is the *empirical content* of such a theory?

Recall that Quine considers the empirical content of a theory to be the set of observation categoricals the theory implies (Quine 1981, 1995, chap. IV). Observation categoricals are compounds of observation sentences which express the fact that the theorist has "become conditioned to associate the one with the other" (Quine 1981, p. 27). In the specific case

[25] We will see later that Quine argues for much the same claim when he introduces the idea of a "pre-established harmony" among our perceptual similarity spaces.

[26] Cf. Grant (2019) for a simple game-theoretic example.

[27] This is what Clark (1996, p. 212ff) calls *joint construals*.

of the child learning the language, the relevant observation categoricals will link linguistic expressions to the situation in which the child saw the linguistic expressions being used, typically the verbal behavior of the people from whom she learned the expressions. Notice that, here, there is no room for private psychological mechanisms, since the child does not have access to her own psychological process, much less to the psychological process of her tutors, and therefore cannot be conditioned to associate a linguistic expression with such psychological processes. As a consequence, these private psychological mechanisms cannot feature in the theory the child is constructing for her language, and, thus, they cannot be part of the meaning of her expressions. Moreover, we have as a direct consequence that anything going beyond the observable behavior of others will also indeterminate, since it will not figure in the relevant observation categoricals and, hence, will not be constrained by the theory's empirical content.[28]

Note that this argument connects the Learnability Argument and the Communication Argument examined above. In learning the language, the child forms observation categoricals after being conditioned to pair observation sentences describing the utterance of certain expressions and the circumstances in which they were uttered. Since this conditioning typically takes place during communication, we have the relevance of communication for the argument. Finally, the argument also avoids the contentious nature of the premises we analyzed above, as it appeals only to what the child can observe, and it is surely platitudinous that the child cannot observe private psychological processes.

Nevertheless, there is still a point of contention, namely, can we assimilate the child to the theoretician? Are children really building theories in learning a language? Obviously, if we construe "building theories" as explicitly and reflectively theorizing, then the answer is negative. But that is not what Quine meant. Rather, Quine is here appealing to his hypothesis concerning language acquisition, according to which, when learning a language, the child is also learning a theory. Consider, for instance, the way Quine accounts for the way the child learns the meaning of a sentence such as "Snow is white". At first, the child becomes conditioned to associate "Snow" with snow and "White" with white objects. But since becoming conditioned to associate "White" with white objects includes being conditioned to associate "White" to snow, this opens the door to becoming conditioned to associate "White" with the *word* "Snow", thus effecting a "transfer of conditioning" (Quine 1973, p. 65).[29] Hence, in learning the meaning of an eternal sentence like "Snow is white", the child is also acquiring a bit of theory, in the form of a disposition to assent to the same sentence when queried.[30]

[28]This reasoning is explicit in Quine (1968/1975, p. 58).

[29]Note that, as Quine (1973, p. 68) observes, this involves a confusion between use and mention.

[30]Note that the person may not even represent the theory explicitly; like Churchland, one can think of a theory here as a "point in [the] individual's synaptic weight space" (Churchland 1990, p. 227). The child

The idea, then, that the child is building a theory can be cashed out by the hypothesis that, in learning a language, the child is also being conditioned to assent to a body of sentences when queried. To say, then, that the child has a theory of meaning (or, more generally, of language), whose empirical content is given by the relevant observation categoricals, is then to say nothing more that the child is acquiring certain dispositions when acquiring a language. Again, since the child cannot be conditioned to respond in any way to private psychological mechanisms, these cannot form part of her theory of meaning (in this sense), and therefore are not part of the meaning of the expressions.

It is important to note that, in contradistinction to the two previous arguments, this line of reasoning does not lead to the conclusion that meanings are public *entities*—in fact, meanings are not entities at all. In other words, a central problem with the two arguments above is that they tried to construe meanings as public entities, as if meanings were *things* that can be grasped by more than one person, in the manner that cups are things that can be grasped by more than one person. Instead, what the present line of reasoning is highlighting is that one is here *reconstructing* a notion of meaning as dispositions to linguistic behavior. The sense in which meanings are public, then, is the sense in which people tend to react similarly in similar linguistic circumstances. This requires no postulation of meanings as entities, public or private:

> Semantics is vitiated by a pernicious mentalism as long as we regard a man's semantics as somehow determinate in his mind beyond what might be implicit in his dispositions to to overt behavior. It is the very facts about meaning, not the entities meant, that must be construed in terms of behavior. (Quine 1969c, p. 27)[31]

This very fact, however, also dooms any attempt at using meanings as explanatory posits. If meanings are understood as regularities in the behavior of speaker's of a given language, then they are not what is doing the explaining; rather, they are what needs to be explained. As Quine repeatedly emphasizes, each individual's history in acquiring a language is different:

> Beneath the uniformity that unites us in communication there is a chaotic personal diversity of connections, and, for each of us, the connections continue to evolve. No two of us learn our language alike, nor, in a sense, does any finish learning it while he lives. (Quine 1960/2013, p. 12)

then "builds a theory" in the sense that a neural network builds a theory, i.e. by having the synaptic weights adjusted accordingly. Cf. Churchland (1989) for more on this idea.

[31] Cf. George (2000, pp. 25ff) for more on this idea.

That, therefore, in spite of such differences, we end up with the same behavioral dispositions is something that cries out for an explanation. If meanings just *are* these behavioral dispositions, then obviously they cannot do the explaining. In the next section, I will turn to an examination of what, in Quine's views, will do the explaining. Surprisingly, the explanatory burden will fall on private psychological states, suitably conceived.

1.3 Theoretical Fruitfulness

In the last section, we saw how Quine considered the publicity of meaning as something to be explained, not as something that does the explaining. It follows that, if meanings are construed as public, then they cannot figure in theoretical fruitful explanations. But why not? Suppose we want to explain why John said "This paper is red" to Smith. One possible explanation is this. John *believed* that Smith wanted to know the color of the paper. He also *desired* that Smith fulfill his want, and, finally, *knew* that the sentence "This paper is red" *means* that the demonstrated paper is red, and that this meaning is public, so that Smith knows that as well. Thus, he uttered the sentence. In this explanation, knowledge of the meaning of the sentence, together with the beliefs and desires of John, are what explains his verbal behavior. The publicity of meaning seems to figure essentially in the explanation, so what is amiss?

The problem, according to Quine, is that this is not an explanation at all—at least, not a *scientific* explanation. Consider an analogous situation. John believes that Smith wants to know the color of the paper, John wants Smith to know the color of the paper, and knows that the sentence "This paper is red" means that the demonstrated paper is red. Yet, he fails to utter the sentence, because he forgot about it, or he has other desires that overrule his desire to let Smith know the color of the paper, or for some reason he cannot bring himself to tell Smith the truth, or any other myriad of cases. In order to account for these types of situations, we insert exceptions in our explanation: if John believes that Smith wants to know the color of the paper, and if he desires that Smith know the color of the paper, and he knows the meaning of "This paper is red", and he does not forget any of these previous conditions, and has no overriding desires, and is not in a state in which he cannot bring himself to tell Smith the truth, etc., then he will utter the sentence in question. More pointedly, one can say that, *ceteris paribus*, if John has the belief and desires in question and knows the meaning of "This paper is red", then he will utter the sentence. But this threatens to render the explanation utterly vacuous, since, without determining the content of the *ceteris paribus* clause in advance, it seems that any supposed counter-example to the explanation will be able to be ruled out as falling inside it.

Now, Quine holds that it is *not* possible to determine the content of this *ceteris plausible*

in advance. That is because this type of explanation, which appeals to the belief-desire psychology of the speaker, is essentially a kind of *projection* from the interpreter into the speaker. That is, in this kind of explanation, "we project ourselves into what, from his remarks and other indications, we imagine the speaker's state of mind to have been, and then we say what, in our language, is natural and relevant for us in the state thus feigned." (Quine 1960/2013, p. 200) This type of methodology "contrasts strikingly with the spirit of the objective science at its most representative" (ibid.), since "what is involved is evaluation, relative to special purposes, of an essentially dramatic act" (ibid.). In other words, for Quine, this type of explanation *cannot* specify in advance the content of the *ceteris paribus* clause, since this content is *relative to the special purposes* to which the explanation at hand is being put, and so will vary according to these purposes. This contrasts with the spirit of objective science, since in the latter the content of such clauses is not so relative, and can be thus specified beforehand.[32]

Notice that the account in terms of belief-desire psychology *appeared* to be explanatory—until we probed deeper into its alleged explanatory credentials. That is why Quine holds any appeal to mentalistic terms suspect: they seem to be explanatory when they are not, or, as Quine puts it, the "easy familiarity of mentalistic talk is not to be trusted" (Quine 1975, p. 95). In light of this, one would expect that Quine would insist that psychological explanations be couched in behavioristic terms. And, indeed, that is what he urges us to do: "Until we can aspire to actual physiological explanation of linguistic activity in physiological terms, the level at which to work is the middle one; that of dispositions to overt behaviour" (ibid.). Importantly, however, that *does not mean* that behavioristic explanations are revealing, for Quine. On the contrary, the last quotation continues: "Its [viz. of explanation in terms of behavioral dispositions] virtue is not that it affords causal explanations but that it is less likely than the mentalistic level to engender an illusion of being more explanatory than it is" (ibid.). In other words, the use of behavioral dispositions in explanations is also held to not be really explanatory. This use is to be commended only because it is less likely to *sound* like a true explanation.[33]

Recall that, according to Quine's official doctrine (Quine 1969b, 1973, pp. 8-15), dis-

[32]Here, it may seem that Quine is adopting what is currently called a *simulationist* theory of mind, in contrast to the *theory-theory* theory of the mind. One could therefore object that, according to the theory-theory theory of the mind, the interpreter is actually developing a full-blown *theory* of the other's intentional states, and is thus much closer to science than Quine allows. I think this is mistaken. Quine can grant the theory-theory theory of the mind, and even grant that this theory has many commonalities with scientific theories, after all he has always argued that science forms a continuum with common sense. Still, science is *refined* common sense, and here the added refinement matters.

[33]So Quine parts company here with Skinner, who did hold that behavioristic explanations were causal explanations. Cf. Verhaegh (2019) for more on the difference between Quine and Skinner.

positional terms are best seen as promissory notes for deeper explanations which reveal the mechanisms underlying the dispositions. So, for instance, in the case of solubility in water, the deeper explanation is in terms of the molecular structure of water and of the soluble substances. Note that there may not be a *single* mechanism underlying the disposition, that is, it is possible that a dispositional term may not mark a single underlying phenomenon, and therefore does not constitute a natural kind.[34] Hence, the important point is not that dispositional terms are natural kind terms (they may very well not be), but rather that dispositional terms are *temporary* checkpoints on the road to a fuller explanation. Once this fuller explanation is attained, they may be eliminated from science.

In the specific case of psychological dispositions, the fuller explanations in question can only be in terms of the organism's structure, and particularly its neurophysiological structure. Unsurprisingly, then, Quine holds that the only purported explanation of linguistic behavior which is truly *scientifically* explanatory is a neurophysiological one, which places the behavior in a causal context. This has two interesting consequences. We saw above that Quine held that behavioristic explanations were superior to mentalistic ones *only* in that they do not engender an illusion of explanation. Conversely, then, as long as we are clear that mentalistic explanations are only placeholders for neurophysiological explanations, we are free to use them. As Quine himself puts it:

> Mentalism thus has its uses as a stimulant. Like other stimulants, it should be used with caution. Mental entities are unobjectionable if conceived as hypothetical physical mechanisms and posited with a view strictly to the systematizing of physical phenomena. They should be posited in the hope of their submitting someday to a full psychological explanation in turn. (Quine 1973, pp. 33–34)

Second, since the speaker's theory of meaning is couched in terms of behavioral dispositions, Quine must hold that this description of the speaker's knowledge will eventually be eliminated in terms of a neurophysiological theory. But this means that underlying the speaker's theory of meaning one will find neurophysiological structures, in other words, *private psychological states*. This opens up an opportunity for a Quinean *reconstruction* of the concept of meaning in terms of private psychological states, as long as we are clear that such states are being considered as proxy for underlying neurophysiological mechanisms. In the next section, I will argue that Quine himself already provided for the beginning of such a program.

[34] Note that Quine (1969b) still spoke of dispositional *states*. He switched to talking to dispositional *terms* only in Quine (1973). The move from states to terms allows him the necessary room for holding that the terms may cover many underlying states. Cf. Hacking (1990) and Quine's reply (Quine 1990a).

1.4 Meanings as Private Psychological States

Let us recap. We saw in section 2 that Quine construed our linguistic practices in terms of behavioral dispositions, and, moreover, that such dispositions were not meant by him to be explanatory; rather, they were to be the target of the explanation. Moreover, we saw in section 3 that the explanation to our linguistic behavior was best accounted by private psychological states, especially in the guise of neurophysiological states. In this section, I will develop this Quinean move in more detail, starting with his account of stimulus meaning and of pre-established harmony.

1.4.1 Stimulus Meaning and Pre-Established Harmony

The explanatory fruitfulness of private psychological mechanisms is clear in the late Quine's account of *stimulus meanings*. This notion was defined in *Word and Object* as being the ordered pair of the affirmative and negative stimulus meanings of a sentence S for a given speaker, where the affirmative stimulus meaning of S for the speaker is the set of all stimulations that would prompt her assent to S if queried (and similarly for the negative stimulus meaning); *stimulations* should be understood here as the triggering of one's sensory receptors.[35] This notion is supposed to be the point of contact between language and the world, thus explaining what is objective about our knowledge.

Initially, the account was supposed to go like this. Obviously, no two different speakers can share stimulations, since their sensory receptors are different. Still, they can be similarly stimulated—or so Quine held at the time of *Word and Object*. This similarity is then the basis for the similarity in the stimulus meaning of their observation sentences, which in its turn is the basis for their intersubjective credentials. Since the objectivity of science depends on the intersubjective credentials of these observation sentences (as a theory's empirical content is given by observation categoricals, which are compounds of observation sentences), it follows that the objectivity of science depends on the similarity of the stimulations for different speakers. This similarity, in its turn, was first explained in terms of an "intersubjective homology or near-homology of nerve endings" (Quine 1995/2016b, p. 35). Hence, the objectivity of science came to rest, for Quine, on the existence of this homology.

Quickly, however, Quine became dissatisfied with this explanation, rejecting the existence of such a homology.[36] This left a gap open in his account, for now he had to explain

[35] As Gary Kemp (2021, p. 10) notes, *stimulation* is here actually ambiguous: it can mean either the *triggering* of one's sensory receptors or the *cause* of such triggering. In his article, Kemp goes on to develop a notion of stimulus meaning, and thus of observation sentence, based on the latter idea, as a substitute for Quine's official notion. Although interesting, Kemp's construction is not relevant for the point I am making here.

[36] Cf., e.g., Quine (1973, pp. 23-4).

why people assented to the same observation categoricals without presupposing sameness of stimulus meanings. One way out, urged by Davidson (1990/2004) and Føllesdal (1995), was to redefine stimulations as being the *causes* of the triggering of one's sensory receptors, instead of the triggering itself.[37] Since there is no problem in saying that the same object causes different triggerings of sensory receptors in different people, there would be no problem, under this definition, in saying that different people could associate the same stimulus meanings with the same sentences. Objectivity would thus be restored.

Quine, however, rejected that way out. First, there are some misgivings about how to construe the causes of the triggerings in certain contexts. In the case of "Lo, a rabbit", the cause may be located in the object itself, but what about "It's raining" or "It's cold"? One could, perhaps, adopt *situations* as being the cause of the triggerings, but Quine (1993, p. 114) finds this hopeless. Second, and more importantly, to adopt such a solution would be to essentially give up Quine's project of an epistemology naturalized. Recall how Quine defined the project in the eponymous essay:

> 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. The relation between the meager input and the torrential output is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology; namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence. (Quine 1969a, pp. 82-3)

In other words, epistemology naturalized, in the sense of Quine, studies, among other things, how subjects form their theories of the world. But, for Quine, this phenomenon is inseparable from how subjects learn to *reify* certain linguistic constructions and thus become able to use language referentially. Thus, part of his project is to explain how we are able to refer to objects, so that the project cannot *assume* at the start that we already refer to objects. If he had located the stimulation at the level of the object, however, he would be assuming precisely that. Quine is quite clear in this regard:

> But I remain unswerved in locating stimulation at the neural input, for my interest is epistemological, however naturalized. I am interested in the flow of

[37] Notice that, again, cause is here ambiguous. On the one hand, it can be the *distal* object, as in Davidson's and Føllesdal's proposals. On the other hand, it can be the *field of forces* that impinge on the sensory receptors, as in Kemp's (2021) proposal.

> evidence from the triggering of the senses to the pronouncements of science; also in the rationale of reification, and in the credentials, if any, of the notion of cognitive meaning. (...) And thus it is that the subject's reification of rabbits and the like is for me decidedly part of the plot, not to be passed over as part of the setting. (Quine 1990b, p. 3)

It is important to be clear here on the function that the notion of stimulus meaning is playing in Quine's hands. Davidson and Føllesdal think that this notion is serving the function of being a reconstruction of the notion of meaning, perhaps a substitute for it. That is, they think this notion is semantical, and they are therefore befuddled by Quine's insistence on locating the notion of stimulation in the triggering of the subject's receptors. If the notion were semantical, then this would be to locate meanings in the privacy of the subject's psychological mechanisms, and would make a mystery of the social character of meaning. Quine, on the other hand, is explicit that this notion is *not* semantical: "The phrase 'stimulus meaning', in other writings of mine, is clearly the villain of the piece. It is not part of the intersubjective business of semantics; I mean it rather in physicalistic parody of the mentalistic notion of private meaning" (Quine 1993, p. 114n1). This is not a side point, but is actually crucial to understand Quine's project.

As I discussed in section 2, Quine considers the "intersubjective business of semantics", and in particular the social character of language, as an *explanandum*, not as an *explanans*. Indeed, Quine's project is to explain the phenomenon of language *without* appeal to intentional notions, which he deems suspect. The notion of stimulus meaning is thus part of this project, and its purpose is to give a physicalistic acceptable account of "the interface between the subject and the external world" (Quine 1993, p. 113), one that does not invoke intentional or mentalistic notions. It accomplishes that by locating the relation between the subject's language and the world in the verbal conditioned responses that the subject acquired as a result of being stimulated in a certain way. The stability of the subject's verbal behavior is then explained by the similarity between society's responses to similar stimulations.

We are thus back at the problem initially outlined. The account sketched in the previous paragraph is able to explain the stability of the subject's verbal behavior, but it is not able to explain the stability of language use across subjects, since there is no tractable notion of similar stimulations across subjects. It is at this point that the notion of pre-established harmony intervenes. Apparently, Quine first came up with this idea in response to Gary Ebbs's (1994) review of *Pursuit of Truth*, and the notion then became central to his late account of the objectivity of observation sentences, particularly in *From Stimulus to Science* (Quine 1995) and "Progress on Two Fronts" (Quine 1996).[38]

[38] Cf. Quine (1995/2016a,b) for his reaction to Ebbs's review. See also his reply to Alexander George (2000)

The idea is simple enough. There is no homology between the sensory receptors of different people, but nevertheless they classify the same situations as being similar to each other. That is, say that two subjects witness two events. Then the subjects's perceptual similarity standards are said to be *in harmony* if, and only if, if the neural intakes of one of the subjects on the two occasions are perceptually similar by her standards, then so are the neural intakes of the other subject on the two occasions also similar by the other subject's standards (Quine 1995/2016a, p. 31). This allows Quine to explain the objectivity of observation sentences as follows. If two subjects speak the same language and their perceptual similarity standards are in harmony, then they will, first, classify the same *sounds* as similar, thus accounting for their shared vocabulary, second, they will classify the same situations as similar, thus accounting for their similar verbal behavior when presented with similar situations.[39]

How is that possible? Quine's answer is that *natural selection* has enforced roughly the same similarity standards to be encoded in each subject's neurophysiology. Quine takes it as a given that an organism's perceptual similarity standards being attuned to the environment is of great survival value. Hence, there is a natural pressure for our perceptual similarity standards to be in harmony with the environment, and, thus, with each other. How this pressure is realized is another matter; here is what Quine has to say on the subject:

> I have no neurological explanation of the harmony, but it would be a matter of a large but limited store of repeatable sensory features of environmental events, lodged deep in the brain of each of us as reusable modules of perception and accessed by each individual's idiosyncratic nerve net. (Quine 2000, p. 408)

Be that as it may, the point is rather clear. We saw how Quine considered the success of our communicative practices, in particular as enshrined in the public character of language, as something to be explained, instead of something that does the explaining. His explanation is eliminativist: he proposes, first, a reduction of the relevant notions to private psychological states *which make no appeal to intentional notions* (stimulus meaning), and, second, an explanation based on natural selection (pre-established harmony) in order to account for the smoothness of our communicative practices. Since, ultimately, the only scientific respectable notions are the neurophysiology of the individuals, there is a sense, then, according to which there are no meanings as public entities, but only private psychological

for the importance he attached to the notion: "George's title 'Quine and Observation' prompted me first of all to scan his eleven pages of footnotes for references to my latest publications; for it is only in them that by a theory of preestablished harmony I dispelled, to my satisfaction, a stubborn thirty-year riddle of the meeting of minds in their perceptions of the shared world." (Quine 2000, p. 408).

[39] There are a couple of wrinkles here. Cf. George (2000, p. 24) for comments.

states that explain why people behave in such a way as to give rise to the *illusion* of meanings as public entities.

1.4.2 Meanings

This results in the following picture. The child is born in a certain initial neurological state, which can be tentatively described in terms of innate learning mechanisms, including, perhaps, an innate learning mechanism that is specific to language acquisition. These mechanisms interact with the stimulus in complex ways, developing into a mature state that account for the person's complex behavioral dispositions.[40] Among these, one finds the person's *linguistic* dispositions, i.e. the neurological states of her brain that underlie her propensities for linguistic behavior. In other words, from a properly *scientific* point of view, the person's language *just is* her underlying neurological state. So why not look into a reconstructed notion of meaning in terms of the person's neurological states as well?

For example, one might construe meanings as being mainly a matter of association, that is, the patterns of node activation in a connectionist network, presumably instantiated in the individual's neurology. In the case of observation sentences, we would expect such nodes to be intimately connected with the individual's sensory receptors. In the case of other sentences, as per Quine's suggestion in Section 1.3 *Word and Object*, the connections may be with other sentences or terms. Perhaps one could even give a more mentalistic description of such meanings along inferentialist lines, hoping to later cash out the inferential connections between expressions in terms of associations in a neural network.[41] Or, more speculatively, perhaps one could construe meanings as the interface between the conceptual and the phonological systems of the subject.[42] Metaphorically speaking, meanings would then be instructions for fetching and building concepts.[43] Here, admittedly, the neurological implementation is not so clear, though a connectionist implementation is not out of question.

[40] Some of Quine's speculations on learning mechanisms, particularly in the opening chapter of *The Roots of Reference*, may mislead one to consider that he subscribes to an "associated chain thesis", namely that the behavioral habits of an organism are best explained by a single "neurological path" starting from the triggering of the sensory receptors and ending in a response. If this were so, then he would be vulnerable to Lashley's (1951) very effective critique of this thesis. But I do not think Quine subscribes to any such thesis. In *The Roots of Reference* itself, Quine (1973, pp. 20ff) already warned about the fact that the organism's own internal states shape its response to stimuli. And in "In Praise of Observation Sentences", Quine (1993, p. 115) also remarks on the non-linear aspect of neural intake. So Quine is also not a traditional behaviorist in these matters.

[41] The compatibility of Quine's ideas with inferentialism is argued for by Kemp (2018), though the suggestion that these might be construed in connectionist terms is mine.

[42] This is the line taken by the Chomskyan tradition. Cf. Jackendoff (1983), Chomsky (1995/2015, 2002, pp. 87ff), McGilvray (1998), and Boeckx (2010, pp. 116ff).

[43] This model is developed in great detail by Pietroski (2018).

Whichever model is adopted, the point of reconstructing meanings as these types of psychological states is explanatory. The initial neurological state of the child explains how she is able to acquire a language, endowing its words with meanings, particular in the cases where her linguistic inputs where greatly impoverished, as in the cases studied by Goldin-Meadow.[44] Later, the *pre-established harmony* among the current states of the language users help to explain how they are able to communicate, even in spite of the differences among such states.[45] So, unlike the public entities supposed by the Publicity Thesis, meanings as private psychological states have a job to do in our current best theories of language acquisition and communication.

There are a couple of important consequences of this move. First, meanings are *internalistically* construed. They are private psychological states of an individual, cashed out in terms of her neurological state, and thus can be in principle specified without explicitly going "outside the head". Moreover, being internal, they are also individual, in the sense of not being part of a "common" language—they are rather part of what Chomsky calls an *I-language*, that is, the current state of the individual's linguistic neural equipment. Second, meanings are dissociated from the notions of reference and truth, that is, from semantics as conceived since Tarski.[46] Successful reference and truthful utterances are *achievements*, which are explained in part by our words meaning what they do. They are thus *what needs to be explained*, not what does the explaining. This is a Quinean point, which is again related to his disagreement with Davidson: the fact that we are able to refer to *rabbits* by using the word "rabbit" is what is partially explained by the notion of stimulus meaning, which is here considered as a private psychological state. Of course, Quine himself associated semantics with reference and truth, but this is a mistake.[47]

Finally, evidence for the meaning of a given word need not be reducible to *behavioral evidence*. Obviously, behavioral evidence is important, but one may use all sorts of consideration in investigating what words mean, including neurological, biological, and psychological.[48] On the old paradigm associated with the Publicity Thesis, meanings were *common knowl-*

[44] This initial neurological state is called *Universal Grammar* in the Chomskyan tradition. As we have seen, Quine has no objection *in principle* to the postulation of such a state, though he may disagree with its exact characterization.

[45] These states are called *I-languages* in the Chomskyan tradition. Again, Quine also admits the existence of these states when he talks about the "chaotic subjective diversity" (Quine 1960/2013, p. 7) which underlies each speaker's linguistic dispositions.

[46] As Pietroski (2018, p. 72n26), citing Burgess (2008), reminds us, Tarski's use of "semantics" to characterize word-world relations, and more specifically as dealing with truth, "was oddly stipulative".

[47] Again, the idea that semantics falls short of truth conditions is a staple of the Chomskyan tradition. Cf. Pietroski (2018) and Collins (2020) for two recent defenses of this idea.

[48] Antony (1997) already defended a similar idea, as had Chomsky. Cf. especially the latter discussion of Kripke on Wittgenstein on following rules in Chomsky (1986, pp. 246ff).

edge. Thus, whatever evidence we had for the meaning of a word needed to be evidence that was *in principle* available to the speakers, since the speakers needed to *know* the meaning of these words. Now, since meanings are dissociated from *knowledge* of meanings, we can appeal to evidence that is *not* in principle available to the speakers.[49]

1.5 Conclusion

One possible line of resistance against the above position comes from certain common-sense considerations. Taking those into account will bring into relief the exact import of the Quinean position I am urging. Quine took meanings to be behavioral dispositions. If two beings are communicating effectively, then that must be because they share the same linguistic dispositions. Their internal psychology seems not only irrelevant, but besides the point. Indeed, if a Martian came to Earth and started speaking English, it seems that the only relevant facts for assessing *that* he is speaking English are behavioral facts.[50] Does he utter strings which are phonetically similar to English ones? Does he react appropriately when queried? As Gregory puts it, the fact that a given community speaks a given language is *constituted* by "the sharing, across a community, of dispositions to linguistic behavior, said dispositions being evidenced by the fluency of discourse within the community" (Gregory 2008, p. 104); thus, given that, by hypothesis, the Martian shares these dispositions with us, then he speaks the same language, and therefore, presumably, means the same things by his words as we do.

Is this the end of the story for Quine, though? Recall that what explains the fact that we share these behavioral dispositions is that there is a pre-established harmony among our perceptual mechanisms. In the case of *humans*, there is some plausibility to that suggestion, since we all have a common ancestry and genetic makeup. But in the case of *Martians*? Perhaps we would come to the incredible discovery that their perceptual similarity standards were the same as ours, but it is much more plausible to suppose that this simply would not be the case. Indeed, if the differences proved to be too great, it would probably put into question the initial impression that they meant the same thing by their words as we did.[51]

Be that as it may, the point is that, even if we grant that the Martians are indeed speaking

[49] Cf Pietroski et al. (2009) for such an investigation into the meaning of "most".

[50] The Martian thought-experiment was used by Katz (1981, p. 89) to argue against Chomsky's psychological approach. A similar consideration is brought up by Kemp (2006, pp. 61-2), also against Chomsky, but this time using a genetic disease.

[51] The case of Kemp cited in the previous note is a bit different. Kemp asks us to consider a genetic disease that altered a person's language processing mechanisms. In that case, we may indeed say that the person means roughly the same thing as we do. But what if the disease affected the person's *perceptual* mechanisms? How far should we allow for deviance in this case? I do not think our intuitions are so clear cut in this other scenario.

English, this would not be the end of the scientific story. Rather, it would be the *beginning*, since we would now have a phenomenon that, again, cries out for explanation: *how* is it that they speak English? And it seems clear that, in order to answer that, we would need to know more both about their psychological states and their evolutionary history. So the appeal to psychological states seems unavoidable.[52]

[52]Collins (2018, pp. 175ff) also takes a similar line in his Chomskyan reply to Katz.

Chapter 2

Lewis's on Codes, Grammar and Communication

2.1 Introduction

Famously, Lewis holds that language is a social phenomenon aimed at producing rational coordination. More specifically, he holds that language's main function is *communication*, where this is understood as the imparting of information about the world (Lewis 1980/1998, p. 22). Since we use our sentences to impart information about the world, such sentences must somehow encode this information; they must represent the world as being a certain way, which endows them with truth-conditions—hence, Lewis's (1970/1983a, p. 190) often cited remark to the effect that "Semantics with no treatment of truth conditions is not semantics". Since, for Lewis, language is primarily a communication device, and communication is a matter of coordination among (rational) agents, Lewis thought he could analyze the use of a language among a given population as being essentially a matter of *conventional* coordination among rational agents. This conventional coordination required two basic attitudes when engaging in the activity of communication, namely, to be truthful in a given language (i.e. to really impart information about the world, using the language in question) and to be trustful in the same language (i.e. to accept that the other is really using the language to impart information about the world).

Lewis's project was, in a sense, a reductive one. He thought that the two basic attitudes briefly described above were sufficient to give a full characterization of language use. In other words, he thought it possible to *reduce* language use to patterns of behavior that conformed to certain regularities in action and belief, that is, patterns that constituted the conventions of being trustful and truthful in the language in question. This reduction was thought to defuse a tension in his account of the nature of languages. On the one hand, languages are abstract objects, functions from sentences to meanings, while, on the other

hand, language is a social activity. His account allowed him to defuse the tension because he could say explicitly how the language, as an abstract object, was *used* by a population, namely if we could describe the patterns of linguistic behavior of the population as conforming to conventions of being trustful and truthful in the language. General semantics would describe the structure of the abstract object, whereas game-theoretical notions, such as coordination problems and his own notion of "scorekeeping in a language game", would describe the social activity itself.

This strict division creates a puzzle. Given the (potential) infinitude of natural languages, in order to interpret another person, I need some way of decoding the meaning of her utterances in terms of the meaning of its parts. Call the competence by which I am able to do so a *grammar*. According to Lewis's story, grammars have nothing to do with language use, since no mention of grammars is needed in order to characterize the relevant conventions that constitute such a use. Of course, as theoreticians, we must employ recursive devices to characterize the abstract object used by a certain population, but this is a limitation of us as theoreticians, not a feature of our linguistic practices. For Lewis, it is at best a contingent empirical fact about us humans that we make use of grammars to communicate, but this fact tells us nothing about the nature of linguistic communication.

This last contention is, to my lights, unacceptable, and it is at least *prima facie* unintuitive. So why did Lewis end up embracing this position? Indeed, as I shall argue in the first section of this paper, even for him this was an unstable position, since he attributes a central place to grammars in settling which language is used by a given population. The answer has to do with the way he conceived our notions of language use and meaning as being *defined* by their theoretical roles in our common sense psychology. Given this conception, he is all but forced to relegate grammars to a secondary position in determining the nature of linguistic meaning, since grammars are not typically taken to be objects of common sense psychology. Still, the resulting position is not entirely stable, and I will argue that it must be ultimately abandoned in favor of one that gives a more central role to grammars.

More specifically, I propose to replace Lewis's notion of a language as an abstract object with an internalist notion of a language as an internalized grammar. As expected, this has ramifications on how to conceive the social activity of language use, since which language is used by a given speaker will not be a conventional matter anymore. Still, Lewis's insight that communication involves a *coordination problem* will be retained, though the nature of this coordination will be changed. Whereas Lewis thought of this coordination as being essentially a signaling activity, here I will tentatively suggest that it is more open ended.

Accordingly, this essay is structured as follows. In the first part, I will explain Lewis's code model of language, according to which sentences (in context) are best viewed as *codes* for pieces of information. In the second section, I will explore Lewis's puzzling position

on grammars, showing how even his austere code model had to assign an important role to them. In the second part, I will examine Lewis's conception of theoretical role and of common sense psychology, and more specifically how he tied the notion of meaning to the notion of truth conditions. In the end, I argue that the position is untenable, and that he ends up appealing precisely to grammars in order to explain how the relevant regularities in behavior obtain, thus violating his own strictures on admissible explanations. In the next section, I will sketch what appears to me to be a more fruitful approach, one that incorporates central insights of Lewis's account, but relocating them inside a more psychologistic context.

2.2 Lewis's Code Model of Language

It will be useful to start with Lewis's rough characterization of a typical communicative situation:

> I know, and you need to know, whether A or B or . . . ; so I say a sentence that I take it to be true-in-English, in its context, and that depends for its truth on whether A or B or . . . ; and thereby, if all goes well, you find out what you needed to know. My choice of what to say is guided by my beliefs. It depends on whether I believe the proposition true at exactly the A-worlds, or the one true at exactly the B-words, or . . . . In the simplest case, the sentence I choose to say is one whose propositional content (in English, in context) is whichever one of these propositions I believe. (Lewis 1980/1998, p. 37)

Although very rough, there are already a couple of features of this model that stand out. First, Lewis considers that the primary function of language is to *impart information*. In the case above, one person knows a piece of information, and the other person needs to know it, and language is one of the best ways of transmitting this piece of knowledge. Second, because the primary function of language is to impart information, he holds that sentences (in contexts) serve to *code* pieces of information. We will see later how to best analyze this, but, for now, suffice it to say that the basic idea is that the participants in a conversation take part in a convention that describes a code for how sentences line up, so to speak, with the way the world is. Redescribing the above situation according to these two ideas, we have the following *code model* of language.[1] One of the participants, the speaker, is in a situation of knowing how the world is. Since he cannot directly transmit the information

[1] This is similar to the code model described by Sperber and Wilson (1986/1995, pp. 3ff), with a couple of differences. The most important is that Sperber and Wilson consider that, in the code model of languages, sentences are codes for *thoughts*, whereas here they are codes for *the way the world is*.

from his mind to the other person's mind, he *encodes* this information in a sentence, so that the other person, upon hearing it, can *decode* it and, so, gather the piece of information she needed. Communicative success in this model is thus achieved if the piece of information you encoded and the piece of information I decoded match.

There are two important consequences of this model. One, since the communicative success is defined solely by the retrieval of the piece of information encoded in a sentence (in context), *how* the encoding and decoding process works is irrelevant, so long as they deliver the sought after result. That is, it is not particularly important that speakers of a language share an encoding or decoding scheme, as long as the schemes agree on the piece of information associated with the sentences in use among them. Obviously, it is to be expected that speakers use the same encoding and decoding scheme, but this is at best an empirical assumption about speakers, not a fact that follows from the model. Two, if we identify *what is said* by a sentence (in context) with the piece of information it encodes, it follows immediately that *what is said* by a sentence (in context) is determined *previously* to the utterance of the sentence. The role of the speaker is to encode this piece of the information and the role of the hearer is to decode it, so that these are *autonomous* actions that, importantly, do not essentially affect the coded message.

Now, for communication to proceed according to a code, people must reliably use the code, i.e. sentences (in context) must reliably stand for ways the world is, and speakers and hearers must know this, so that they can use the sentences (in context) to retrieve the piece of information they are after. This reliability implies *regularity*, that is, the use of language must be regular enough for people to form expectations concerning this use. We are thus naturally led to Lewis's conventional account of language use. Let us explore this idea a bit further.

We can start with Lewis's definition of a language $\mathscr{L}$ as a set of ordered pairs from sentences to meanings, thought of here as ways the world could be; this definition is perfectly natural given the coded model of language, as, according to this definition, what defines a language is precisely that its sentences stand for, or encode, pieces of information. The next step is to define the important notion of how such a language is *used by* a population P. The definition is rather intricate, but the core point is simple enough. Say that a person is *truthful* in $\mathscr{L}$ iff she avoids uttering false sentences of $\mathscr{L}$, and say that a person is *trustful* in $\mathscr{L}$ iff she expects others to be truthful in $\mathscr{L}$. A language $\mathscr{L}$ is used by a population P iff there is a convention among P of being truthful and trustful in $\mathscr{L}$.[2] Note that these conventions of truthfulness and trustfulness have as an effect that people can reliably expect

[2]I am skirting over many details here, and in particular the controversy regarding Lewis's account of conventions, since they will not matter for the point I want to make. For those interested, cf., aside from Lewis (1969, 1975/1983) himself, the survey by Lepore and Stone (2015a) and the references therein.

specific sentences to stand for specific pieces of information, that is, they can reliably expect others to be using the same *code* as they are. We thus see how the main features of Lewis's account of language actually flow rather naturally from the code model.

There are three points worth emphasizing about this idea. First, a language is here defined as an *abstract object*, indeed, an *extensional* one. Again, this entirely natural under the code model, since a code *is*, essentially, an abstract object, precisely a function from codes or signals to encoded messages. Second, as mentioned above, it is the associated piece of information that is essential for this model, so that the encoding and decoding schemes are peripheral to the main function of a language, something that is reflected in the fact that languages as thus defined *directly* pair sentences with meanings, thus bypassing the compositional structure of sentences. Finally, since the utterances of a person will receive their meaning from the language which is conventionally adopted by the population to which that person belongs, it follows that the meaning of a person's utterances are grounded on those social conventions. That is, as the reliability of the code presupposes that people regularly use it adequately, a person's utterances must conform to the reliability expectations of other users of the code. There is thus very little room, according to this model, for novel uses of language.

Let us focus a bit on the second consequence mentioned in the last paragraph, viz. that since sentences (in context) are directly paired with pieces of information, the procedures for encoding and decoding sentences have only a marginal role in the account of language and language use. Of course, as noted in the beginning of the section, it may be a psychological fact about humans that they employ specific procedures both in speech production and in speech comprehension, and even that these largely match, but it is conceivable, on Lewis's account, that a given population could use a language without having any such internalized procedures. To make this vivid, imagine a language with only a finite number of sentences (perhaps some kind of signaling language, as analyzed by Lewis (1969, chap. IV)). In that case, there would be no need for a grammar, since the users of the language could directly read the piece of information encoded by sentences from the sentences themselves (say, by consulting a table). Indeed, as Sperber and Wilson note, this is a "reasonably good description of what often happens when artificial codes are devised and used" (Sperber and Wilson 1986/1995, p. 26)

Natural languages, however, are notoriously *not* like that, i.e. they have an *unbounded* number of sentences. Lewis is of course aware of that, so that his languages are typically thought of as infinite. In the case of a language with an infinite number of sentences, it is not humanly possible to construct a table which directly gives the pairs of sentences and pieces of information. How, then, can we hope to describe such languages? We need to employ *some* method of derivation, that is, some method that allows us to describe an

infinite number of sentences with an infinite number of pieces of information *on a finitary basis*. The most natural method here would be to start with some basic clauses and apply recursion to obtain the meaning of the derived clauses. That is, the most natural method is to consider a *compositional grammar* which assigns a complex structure to sentences and makes the meaning of a sentence depend on this structure and the sentence's constituents. Hence, grammars appear to be useful merely in achieving conciseness: "A concise grammar for a big language—for instance, a finite grammar for an infinite language like ours—had better work on the compositional principle" (Lewis 1980/1998, p. 25).

Thus, according to Lewis, the need for grammars is then a *parochial* fact about us humans *in our theoretical attempt to describe a language*. It is not a fact about language use as *such*, i.e. it is not a metaphysically charged fact about languages. Obviously, this means that, in order to *theorize* about languages, we need to make recourse to grammars, but Lewis thought that this was a fact about our *theorizing*, not about the languages themselves. In other words, Lewis considered grammars as important tools *not* to explain the productivity of language users, but rather as tools for the *theoretician* to build his "systematic restatement of our common knowledge about our practices of linguistic communities".[3]

Of course, this has immediate implications for the psychology of speech production and comprehension.[4] Given that the speakers we are typically acquainted with are humans, and thus finite, it follows that whatever encoding and decoding scheme they are employing must also work from a finitary basis and then use this basis to project the pieces of information to be recovered from arbitrary sentences. Let us suppose, as seems reasonable, that each speaker's encoding and decoding schemes are largely the same,[5] and let us call this encoding-decoding scheme an *internalized grammar*. Here, there emerges a problem. As we saw above, the code model depends on there being a certain regularity in the use of sentences, so that people can form expectations about the reliability of the code. Hence, since they all use internalized grammars for encoding and decoding messages, these internalized grammars must be aligned at least at the level of output, for otherwise there would be no such regularity and no such expectation of reliability. Such an alignment is not a brute fact, but something that cries out for explanation. How, then, do we explain it?

One way would be if such internalized grammars were acquired by a process of social

[3]Cf. Yalcin (2014, pp. 39-43) for a similar view; I was draw to the significance of the quotations in this paragraph by this paper.

[4]This is a huge topic, which I unfortunately do not have the space to explore here. For an excellent overview, cf. Townsend and Bever (2001, Chap. 4).

[5]It seems reasonable, and for the most part it is even true, but in many cases these two may be distinct, as when a person can understand a language without speaking it. Begby (2017, p. 707) calls this alignment between encoding and decoding schemes *bidirectionality*, showing how this property is important by exploring the consequences of its failure.

conditioning. In this case, as the individual would acquire the grammar from the community of speakers to which she belonged, this grammar would automatically align with the grammar of those from that community. Another way would be if speakers were each born with an internalized grammar, and this grammar developed in harmony in each member of the community—perhaps as a consequence of a similar natural endowment, together with a similar social environment. Note, however, that these two accounts have different explanatory directions. The first account explains the alignment between the individuals as a matter of social forces molding their internalized grammars into shape, so to speak. The second account explains the alignment between the individuals as a matter of their natural endowment and developmental path, so that, at the limit, it is the congruence of the internalized grammars that explains the regularity of the language, and not vice-versa.

I emphasize this because Lewis's account opens up a gap between the output generated by each particular individual and the general code that prevails in the speaking community to which they belong. As we will see, this gap will put enough pressure in his account for him to modify it, giving a more prominent role to internalized grammars. This greater prominence, in its turn, will make the second account sketched above more plausible, thus suggesting that Lewis's got the explanatory direction wrong.

2.3 Lewis's Expanded Account: The Role of Grammars

We saw in the last section that Lewis essentially saw languages according to a code model, which relegated grammars to a rather marginal account in the empirical psychology of language use. Later, however, Lewis (1992/2000) would complicate this story in his reply to certain objections raised by Schiffer (1987, pp. 255-261) and Hawthorne (1990), and the added complication is revealing of his general outlook. The problem is this. Recall that a language $\mathscr{L}$ is used by a population P iff members of P avoid uttering false sentences of $\mathscr{L}$ and expect other members of P to do so as well. Suppose $\mathscr{L}$ is infinite, in that there is no bound for the complexity of its sentences. Since there is presumably a bound for the processing power (not to mention the lifespan) of humans, there will be infinitely many sentences of $\mathscr{L}$ that will be impossible to utter—indeed, the majority of them. Let $\mathscr{L}'$ be a language which agrees with $\mathscr{L}$ in its utterable part and disagrees on the rest. Since $\mathscr{L}'$ disagrees with $\mathscr{L}$ only on the part that is impossible to utter, it follows that members of P will automatically avoid uttering those sentences, and will therefore, *a fortiori*, avoid uttering those sentences which also happen to be false. Moreover, they can also reasonably expect others not to utter them, since they are by hypothesis impossible to utter. Hence, there is a convention among members of P of being truthful and trustful in $\mathscr{L}'$, whence $\mathscr{L}'$ is *also* used by P.

This is a problem because facts about the meanings of sentences were supposed to be determined by social facts about the conventions prevailing in a given population. But given that both $\mathscr{L}$ and $\mathscr{L}'$ can be used by a given population, and $\mathscr{L}$ and $\mathscr{L}'$ differ in their attribution of meaning to infinitely many sentences, it seems that facts about meanings of sentences are not determined by those conventions after all. To solve this, Lewis (1992/2000), surprisingly, turns to grammars. The idea is to argue that determination of meaning is a two step process. First, the conventions among a given population fix a given set of languages, all of which agree on the sentences which are possible to utter. We then construct a grammar for this part, and *extrapolate* the grammar to the rest of the sentences. The language generated by this grammar is then selected as *the* language spoken by the population, and so the problem of meaning is thereby settled.

All of this is good, except that it seems to only push the problem back to grammars. By gerrymandering the original grammar, it is possible to construct an alternative grammar which agrees with the first one except for the sentences that are not possible to utter. So how do we know which grammar is *the* grammar used by a given population? One idea, proposed by Loar (1981, pp. 259ff), is that a grammar is the grammar used by a given population if it is the grammar which figures in the correct psycholinguistic explanation of how members of that population understand their language. But then, nothing guarantees that the same grammar will be used by the whole population; moreover, facts about meanings would now be determined by facts about the psychological states of individuals as described by the correct psycholinguistic explanation of how these individuals understand their language, so that we would be back with Meaning Internalism.

Unsurprisingly, then, Lewis rejects Loar's proposal, arguing instead that, among the candidate grammars, some are more *natural* than others, and we "have no difficulty in telling the difference" (Lewis 1992/2000, p. 150) between natural and unnatural grammars. Given this distinction, Lewis then insists that "[w]e can reasonably hope that all [natural] grammars that agree on the used fragment will agree everywhere" (Lewis 1992/2000, p. 151), i.e. that there is in fact only one language generated by the natural grammars. The notion of *naturalness* in play here refers back to Lewis (1983/2000), in which he proposes to consider it as a primitive notion, a move which supposedly helps to explain many metaphysical puzzles. I will not here question this metaphysical move, but I do wish to question its pertinence to the topic at hand.

Grant that naturalness is a primitive notion, and grant that the natural grammars do offer a solution to the above problem. Now, presumably, linguists have their own criteria for ranking grammars.[6] This seems to land us in a dilemma. Either the criteria of linguists align itself with the naturalness criteria or not. If they do align, then the appeal to naturalness

[6]Cf. Chomsky (1965, Chap. 1) for a famous discussion of this problem.

seem superfluous. If they do not align, the appeal to naturalness seems wrong. So why not appeal to *those* criteria, instead of to this mysterious naturalness property? In commenting the possibility of a conflict between the grammar selected by the psycholinguist and one selected based on naturalness, Lewis says:

> Maybe there is a grammar somehow written into the brain. And conceivably it is a[n unnatural] grammar, so that the language it generates differs, somewhere outside the used fragment, from the language we get by straight extrapolation. Schiffer has asked: does straight extrapolation give the right answers even then? I think so. If not, then whenever we resort to extrapolation to answer questions of syntax and semantics, we are engaged in risky speculation about the secret workings of the brain. That seems wrong. (Lewis 1992/2000, p. 151n6)

But why does it "seem wrong"? That is precisely how many linguists—Chomsky in particular—conceive of the matter, i.e. that when we are engaged in questions about syntax and semantics, we *are* engaged in "risky speculation about the secret workings of the brain". Notice the oddity of Lewis's position: it may be that each member of a population P has internalized an unnatural grammar G, and thus that each of them interprets the others according to language $\mathscr{L}$, yet, given the naturalness criteria, they are all actually employing grammar G', which means that they are in fact using language $\mathscr{L}'$. So the language they use to interpret each other's utterances is not the language they actually speak. This seems bizarre. Something has gone wrong. Why does Lewis insist in maintaining Meaning Externalism when, given the way he outlined his own position, the Internalist solution is much more attractive? I believe the answer resides in the way he conceives of his project:

> (...) I am ready enough to believe in internally represented grammars. But I am much less certain that there are internally represented grammars than I am that languages are used by populations; and I think it makes sense to say that languages might be used by populations even if there were no internally represented grammars. I can tentatively agree that $\mathscr{L}$ is used by P if and only if everyone in P possesses an internal representation of a grammar for $\mathscr{L}$, if that is offered as a scientific hypothesis. But I cannot accept it as any sort of analysis of "$\mathscr{L}$ is used by P", since the analysandum clearly could be true although the analysans was false. (Lewis 1975/1983, p. 178)

In other words, Lewis's rejects the appeal to internalized grammars because it is not a good analysis of "$\mathscr{L}$ is used by P". This raises two questions: first, what does Lewis mean by analysis and, second, why is meaning internalism not up for this task? I will turn to these questions in the next section.

2.4 Lewis on Common Sense Psychology

Lewis locates the concept of *meaning* inside a general theory of persons, which constrains how we interpret each other. That is, we each try to make sense of each other by attributing to one another beliefs and desires, as well as attributing meanings to each other's utterances. These attributions form *particular* theory of persons (the theory of Karl, of David, etc.), which are in their turn constrained by our *general* theory of persons, that is, the general theory that relates the concepts of belief, desire, and meaning to each other. Call this general theory of persons *common sense psychology*, since it is the psychological theory the common folk employ in making sense of each other.

One important point about this common sense psychology is that Lewis takes it to be a rudimentary form of scientific theory, with belief, desire, and meaning appearing as its *theoretical terms*. Since this will be important in what follows, let us pause for a bit to see how Lewis understands the notion of a theoretical term.

2.4.1 The Functional Character of Theoretical Terms

Like Carnap (1956) before him, Lewis accepts that scientific theories normally come equipped with a distinguished set of terms, namely its theoretical terms. Unlike Carnap, however, he does not assimilate those terms to terms denoting unobservable entities, so that they would contrast with terms denoting observables. Rather, he contrasts the theoretical terms with the *original terms*, the "*old* term we already understood before the new theory T with its new T-terms was proposed" (Lewis 1970/1983b, p. 79). The picture seems to be roughly this. We are faced with a given phenomenon that we want to explain. This phenomenon is described in a language which we already understand, with its own set of terms, which we may designate the *original* terms. In order to explain this phenomenon, we introduce a theory, T, with its own particular vocabulary, the theoretical terms of T.

In the case of common sense psychology, what is to explained is, according to Lewis, the physical facts about people, in particular their behavior. And what does the explaining is the general theory of persons, especially as couched in the rudimentary theoretical vocabulary of beliefs, desires, and meanings. In other words, Lewis believes that the notions of belief, desire, and meaning are *theoretical notions* introduced to explain the behavior of people, as described in physical terms. So the *original* vocabulary is here the *physical* vocabulary, and the *theoretical vocabulary* is the vocabulary of beliefs, desires, and meanings. As we will see, this is important, because Lewis believes there is a special relationship between these two vocabularies which imprints a certain character to the notions characterized by the theory.

Following Carnap (1974/1995, chap. 28), Lewis believes that, for any scientific theory, it is possible to separate its *empirical* content from its *analytic* content. The idea is simple

enough. Let T be a scientific theory, which we may take to be the conjunction of all its axioms.[7] In the case at hand, we can think of a sentence, $CP(b, d, m)$, as axiomatizing common sense psychology, with constant terms for belief, desire, and meaning. That is, $CP(b, d, m)$ basically spells out our general theory of persons, making explicit the interconnections between belief, desire, and meaning.

Define now the *Ramsey* sentence of the theory, ${}^{R}T$, to be result of replacing all theoretical terms by variables and existentially quantifying the result. This sentence is taken by Carnap to express the *empirical* content of T, as it basically claims that there are entities satisfying the relations and properties postulated by T. In our example, this would be $\exists x \exists y \exists z CP(x, y, z)$, a sentence that asserts that there are such states—viz. belief, desire, and meaning—which satisfy the constraints of our theory.

Finally, define the *Carnap* sentence of T, ${}^{C}T$, to be ${}^{R}T \rightarrow T$, i.e. the material conditional whose antecedent is the Ramsey sentence of the theory and the consequent is the theory itself. In the case of common sense psychology, this would be $\exists x \exists y \exists z CP(x, y, z) \rightarrow CP(b, d, m)$. This sentence basically says that *if* there are states satisfying the constraints of the theory, these states are belief, desire, and meaning. According to Carnap, this sentence expresses the *analytical* content of T:

> It can be easily shown that [the Carnap sentence of a theory] is factually empty. It tells nothing about the world. All the factual content is in the sentence F_T, which is the Ramsey sentence ${}^{R}T$. The [Carnap sentence] simply asserts that, *if* the Ramsey sentence is true, we must then understand the theoretical terms in such a way that the entire theory is true. It is a purely analytic sentence, because its semantic truth is based on the meanings intended for the theoretical terms. (Carnap 1974/1995, p. 279)

Now comes the crucial part. Using this idea, Lewis (1970/1983b) shows how one can explicitly define the theoretical terms of T. The procedure is simple enough: supposing that each theory has a unique realization, one can define a theoretical term τ to be *the* entity that, if the theory is true, realizes the theoretical role assigned to it by the Ramsey sentence of the theory. So, belief, for instance, can be defined as follows:

$$b = \imath y_1 \exists y_2 \exists y_3 \forall x_1 \forall x_2 \forall x_3 (CP(x_1, x_2, x_3) \leftrightarrow y_1 = x_1 \wedge y_2 = x_2 \wedge y_3 = x_3)$$

This may seem like a baroque digression, but it actually goes to the heart of the matter. As I mentioned above, Lewis takes common sense psychology to be a kind of rudimentary scientific theory, with belief, desire, and meaning taking on the role of theoretical terms

[7] I am here assuming, for simplicity, that theory is finitely axiomatizable. If not, however, Lewis is comfortable with invoking infinitary resources, such as infinite conjunctions: cf. Lewis (1970/1983b, p. 80).

in this theory. This means that, as per above, they are, in his view, *defined* by the role they play in this theory. Thus, to investigate their nature simply *is* to investigate the way they figure in common sense psychology. In other words, their nature is *exhausted* by their role in common sense psychology, in such a way that anything outside the purview of this rudimentary theory is irrelevant to understanding them, including any "esoteric scientific findings":

> The concepts of belief, desire, and meaning are common property. The theory that implicitly defines them had better be common property too. It must amount to nothing more than a mass of platitudes of common sense, though these may be reorganized in perspicuous and unfamiliar ways. Esoteric scientific findings that go beyond common sense must be kept out, on pain of changing the subject. (Lewis 1974/1983, pp. 111–2)

Let us recap. We started with a puzzle: why did Lewis think that an appeal to, say, neuropsychology in selecting grammars was illegitimate? In order to answer that, we analyzed how he conceived of the central notion of *meaning*, in particular how it figured as a *theoretical term* in common sense psychology. That is, Lewis thought that the notion of *meaning*, together with the notions of belief and desire, was introduced to help explain the physical behavior of people in terms of a rudimentary theory, namely common sense psychology. Since these notions were introduced in the context of this rudimentary theory, they are *defined* in terms of the functional role they fulfill in that theory, as per the procedure outlined above. It is clear, then, why internalized grammars are out of bounds: the very notion of such a grammar is an artifact of current linguistic theory, and therefore *not* part of common sense psychology. It thus cannot be what determines the *nature* of the concept of meaning, and so appeals to it must be avoided.

One of Lewis's main goals, then, was to provide an *analysis* of those notions, that is, an account which makes explicit the common sense psychology that determines their role. Since this will be important in what follows, it is to this task that I now turn.

2.4.2 Common Sense Psychology and Truth Conditions

In "General Semantics", Lewis remarks that "Semantics with no treatment of truth conditions is not semantics" (Lewis 1970/1983a, p. 190). While some (Pietroski 2018, esp. chap. 2) have forcefully—and, in my view, correctly—argued against this idea, I believe it is nevertheless important to understand *why* Lewis adopted this approach. In this section, I will argue that the equation of semantics with *truth-conditional* semantics follows directly from his assumption that the notion of meaning is defined by its theoretical role in common sense

psychology. This will later, in the next section, allow us to untangle the core of truth of Lewis's idea from the misguided assimilation of semantics to truth-conditional semantics.

Recall that Lewis's conceives of language according to the code model, that is, according to a model that sees sentences (in context) as codes for pieces of information, and the role of the speakers is limited to using those sentences to encode and decode such pieces of information. This model, in its turn, finds its theoretical motivation in a deeper picture of the role of language in our lives. In particular, to understand Lewis's position, it may be helpful to begin by noting that he sees language as a type of solution to a *coordination* problem. A coordination problem arises when two or more agents "have a common interest in all doing the same one of several alternative actions" (Lewis 1969, p. 24). So, to use a well-worn example, suppose two people, Jack and Jane, want to meet. They can meet at either a Chinese or at a Mexican restaurant. So, if Jack goes to the Chinese restaurant and Jane to the Mexican one (or vice-versa), they will not meet and thus be frustrated. To avoid this outcome, they must *coordinate* their actions so that the two go to the same restaurant. Note that both may have no preference regarding which restaurant to go, in which case they must choose *solely* based on their expectation of where the other will go. That is, since there is no forced choice, the relevant factor in choosing where to go is what the other has also chosen. If, whenever we need to meet, we happen to always settle for one restaurant over the other *because* we expect to find the other there, then this solution to the coordination problem is said to establish a *convention* for Lewis. The relevant features of a convention are: (a) it is a solution to a coordination problem, (b) there were other equally good solutions for the coordination problem, and (c) we repeatedly choose this solution because of our expectations about the other agent's expectations.[8]

One important type of coordination problem involves *signaling*. The idea is roughly the following. Among a certain group of people, what action to follow depends on whether a given state of affairs obtains. Not everyone is privy to whether such state of affairs does in fact obtain, though; typically, only one among them is in position to tell that it does obtain. So they come up with the following solution: the person who is in position to tell whether or not the state of affairs obtains performs a given action depending on whether or not the state of affairs obtains. The action thus functions as a kind of *signal* to the others about what has transpired. For instance, suppose a group of people are foraging in an area known for its predators. It is difficult to forage and watch for predators at the same time, so they set up a watcher who will make a loud noise if she spots a predator, or else remain silent. The noise or the silence serve then as signals for the presence or absence of predators, respectively.

[8]Here and elsewhere, I am skirting over a lot of technical details, since they are not particularly relevant for my argument. The interested reader should definitely check out Lewis (1969) for his detailed application of game-theory to the analysis of conventions.

This type of coordination problem, involving signaling, has a couple of important features. First, it is a solution to the coordination problem of how to impart information about potential predators. Second, it is clear that there is a certain arbitrariness to which specific action serves as a signal. Obviously, it must be an action which the audience can perceive, but, within that range, any number of actions would suffice. Finally, the solution is sustained by the expectations of the agents involved: the people foraging *trust* that the watcher will give the signal if, and only if, she spots a predator, and the watcher, on her turn, *does* give the signal if, and only if, she spots a predator—i.e., she is *truthful* in her signaling. These expectations translate into specific *beliefs* and *desires*, e.g., a desire to give the signal only if a lion is present, a belief that the audience also desires a signal to be given only if a lion is present, etc. What this means is that this specific arrangement is a *convention* in Lewis's terms and, moreover, that such a convention is sustained by a specific pattern of beliefs and desires which specify the *function* of the signaling.

The picture that thus emerges is this. There is a tight connection between the beliefs and desires that sustain the signaling convention, on the one hand, and the function of the signal, on the other. If the agents believe and desire that the signal is to be given in the presence of predators, then the signal's function *is* to be emitted in the presence of predators. Conversely, if the function of the signal is to be given in the presence of predators, then we can be certain that the agents will believe and desire that it be given in the presence of predators. In this way, there is an expectation that the signal will be *used* only to indicate the presence of predators—that is, that the *meaning* of the signal is that there are predators nearby.[9] In other words, the fact that the agents believe and desire that a given signal be used to indicate a given state of affairs is taken by Lewis to *constitute* the fact that the meaning of the signal *is* to indicate the state of affairs. This is unsurprising, since, as we saw in the last section, Lewis believed that common sense psychology is a rudimentary theory whose interrelations between the notions of belief, desire, and meaning *define* these notions. Here we see precisely such an interrelation in action, in the fact that we can go either from the desires and beliefs of a population to the meaning of signals or from the meaning of the signal to the desires and beliefs of the population.

One important feature of the above situation is that the agents involved have a very specific situation which they need to keep track of, so they invent a specific signal to represent *that specific, recurrent situation*. Here, it makes sense to speak of a code, since the regularity of

[9]One could also take it to *mean* that one must stop foraging and flee. As Lewis (1969, p. 144) makes clear, in this case it is possible to hold that the signal means *both* that there is a predator and that one must flee, i.e. it is both a signal *that s* and a signal *to do r*—this type of primitive signal with a double function is especially important for Millikan, who calls this type of representation "pushmy-pullyu representations" (Millikan 1995/2005). Here, however, I will concentrate mainly on the indicative function, since it is that function which is central to Lewis's account.

the situation makes possible for the agents to form expectations about it and about the kind of signals used to encode it, so that a convention can take hold. Whether or not Lewis's analysis of language use in terms of conventions can be sustained—and thus the code model for language—depends, then, on whether or not there are sufficiently robust regularities in language use to sustain such conventions. In the next section, I will argue that there are not. Curiously, Lewis's strategy for dealing with this recalcitrant feature of language use again invokes internalized grammars, so that what was once marginal is seen to have an ever increasing role in his account of language.

2.4.3 Languages and Grammars Again

Of course, Lewis is the first to recognize that linguistic utterances differ in many ways from the simple signaling conventions analyzed above (Lewis 1969, pp. 160-1). Most of the differences, however, are only of added complexity, in such a way that the core connection between the beliefs and desires of a population and the function of language remain. We can therefore see why he considered meanings to be necessarily tied to the specification of truth conditions: in this picture, the beliefs and desires which sustain the conventions around the use of a language *just are* beliefs and desires about the truth conditions of various utterances. Hence why the code model appealed so much to him: according to this model, sentences (in context) are precisely *codes* for truth conditions. We therefore seem too have a vindication of Lewis's claim that the notion of meaning is specified inside a general theory of persons, and that this specification involves truth conditions in such a way that "semantics without a treatment of truth conditions is not semantics". Still, there is one blatant difference between signaling conventions and natural languages that threatens the whole project.

Signaling conventions are rigid, in the sense that, once an action is settled upon as a signal, it is important that the person responsible for the signaling sticks to that action. There is no room for creativity or improvisation, since the state of affairs corresponding to the signal must be unambiguously recovered by the others so that they can settle upon an appropriate course of action. There is then a very clear regularity between the use of *specific* signals and the states of affairs corresponding to them. The natural language analogues of those signals, namely sentences, are nothing like that, however. It is a truism that, aside from an insignificant portion, most sentences that people hear are novel. As Chomsky (1959) famously noted, this seems to doom any account of language based on past regularities, since, in the case of language, there are appear to be no such regularities:

> We constantly read and hear new sequences of words, recognize them as sentences, and understand them. It is easy to show that the new events that we accept and understand as sentences are not related to those with which we are

> familiar by any simple notion of formal (or semantic or statistical) similarity or identity of grammatical frame. Talk of generalization in this case is entirely pointless and empty. (Chomsky 1959, p. 56)

Lewis, evidently, agrees. His strategy to accommodating this observation is similar to the one proposed above regarding the unutterable sentences of a language, and is also, in many ways, reminiscent of Quine's (1960/2013) position. A speaker of a given language starts by understanding sentences and forming the relevant expectations with regards to these sentences. Later, she internalizes a grammar that allows her to deal with novel sentences:

> It may happen that a hearer, say, has never before encountered the sentence now addressed to him; but he forms the appropriate belief on hearing it—one such that he has responded trustingly in $\mathscr{L}$—because his past experience with truthfulness in $\mathscr{L}$ has involved many sentences grammatically related to this one. (Lewis 1975/1983, p. 168)

Unfortunately, the same problem we saw above with regards to the choice of grammar reappears here. In order for a speaker of the language to form the relevant belief, the grammar she has internalized must generate precisely the language $\mathscr{L}$. Yet how do we know that she internalized a grammar that generates $\mathscr{L}$, instead of, say, $\mathscr{L}'$, which is identical to $\mathscr{L}$ up to a certain time slice of the speaker's life, but differs afterwards? Notice that here the problem is indeed worse, for it makes no sense to appeal to naturalness in *this* context, since here it is not a matter of the *theoretician's* choice among rival accounts of the data, but of the *language user's* internalization of the correct grammar. The difference is that, in the former case, the theoretician can make *explicit* appeal to naturalness in choosing among rival accounts, whereas, in the latter, the language user, presumably, does not make such an *explicit* appeal—in fact, in this context, naturalness appears to be exactly the sort of "esoteric" notion previously banned by Lewis.

One solution would be to say that the language user makes *implicit* use of naturalness in choosing the relevant grammar. But the only way the language could make implicit use of naturalness in choosing the relevant grammar is if somehow this naturalness constrains her language acquisition process. That is, for naturalness to implicit guide the language user means that the developmental path of the language user in her language acquisition evolves along certain natural lines. This, again, collapses into an appeal to "esoteric scientific findings", since it is an empirical question whether language acquisition does follow along such lines. In the end, then, an appeal to "esoteric scientific findings" seems unavoidable. But if that is so, the whole project of giving an account of meaning in terms of its role in common sense psychology must be abandoned.

This section started by trying to understand how Lewis conceived of language, in particular his idea that the notion of meaning is defined by its theoretical role in a more general theory about the attribution of beliefs and desires to others, a theory he called *common sense psychology*. We saw how this role was connected to the patterns of belief and desire which sustain the conventions of being trustful and truthful in a certain language, as modeled by the simpler case of signaling conventions. Finally, we also saw how, in order to cope with the added complexities of natural language, Lewis appealed to grammars, but that such an appeal eventually resulted in the project's demise. In the next section, I will delve deeper into the reasons for this failure, by considering more closely how communication can be thought of in terms of a rational activity.

2.5 Reassessing Communication and Truth Conditions

We saw in the last section how Lewis's thought of language as essentially a tool for signaling to others about certain states of affairs. This meant that he considered that, in interpreting others, the language users make use of two broad types of knowledge: (1) their expectations about the others' and their own beliefs and desires regarding the communicative situation, with these expectations following a *rational* pattern and (2) grammatical knowledge, which is used to parse novel utterances, in such a way that grammar is an important constraint in interpreting another. I believe that he was right in the two counts, but that, contrary to what he thought, the reasoning supporting (1) actually undermines his thesis according to which the grammar directly delivers truth conditions. Grammatical rules *constrain* truth conditions, but do not fully determine it.

In this section, I will analyze these claims in depth, by first analyzing Lewis's conception of communication and, second, developing a conception of grammar as furnishing a set of *structural* constrains on content, constraints which fall short of truth conditions.

2.5.1 The Rationality of Language Games

We saw in the last section how Lewis's thought of communication as being a collective enterprise, and as such subject to rational principles as outlined, e.g., in game theory. Later, Lewis (1979/1983) would refine this idea to include an account of conversational *dynamics*, that is, a rational account of how conversations *evolve*. In order to give an account of what is specific to Lewis's idea, it will be helpful to contrast it with Stalnaker's similar theory.

Stalnaker (1970/1999, 1974/1999, 1978/1999) had already proposed, working with presuppositions, that communication does not take place in a vacuum. Rather, communication takes place against a background of common assumptions, pieces of information that those communicating take for granted for the purposes of that conversation—what Stal-

naker termed *the common ground*. Notice, first, that this common ground need not be factual, but merely *assumed* to be factual by the participants (they may be mistaken in their presuppositions, or simply not care about their falsity for the purposes of their particular conversation). Second, this common ground is not just a static store of information, but is a store of information that both *constrains* the flow of conversation and also *is updated* by that very flow. So, for example, if I utter the sentence "He is very tall" in a context in which my cousin is salient—in which the fact that my cousin is the subject of the conversation is part of the common ground—, then I will be taken to be naturally referring to my cousin.[10] So the fact that the common ground contained the information that my cousin was salient *constrained* the reference assigned to the pronoun by my utterance. On the other hand, that same utterance will also update the common ground, by adding the information that my cousin is very tall to our shared background.

This last feature of the common ground is important, since it allows Stalnaker (1978/1999) to give a characterization of the speech act of assertion in terms of its effects on the common ground. According to Stalnaker (1978/1999, p. 86), the point, the *goal* of making an assertion is essentially to add a piece of information to the common ground. Whether or not this particular characterization of assertion is correct is not what is here relevant;[11] what *is* relevant is that this characterization points in the direction of a characterization of the various speech acts in terms both of the way they are constrained by and the way they update the common ground:

> Each of these kinds of linguistic action [e.g. making stipulations, temporary assumptions, or promises, asking questions, and giving commands and permissions] is presumably performed against a background of presuppositions, and can be understood partly in terms of the effect that it has, or is intended to have, on the presuppositions, and on the subsequent behavior, of the other participants in the conversation. (Stalnaker 1978/1999, p. 88)

Notice that, although the emphasis here is clearly linguistic, there is nothing that is specific to linguistic practice in the idea of a common ground. On the contrary, as Stalnaker emphasized in later writings, the notion of a common ground is in play in any kind of collective activity: "Any cooperative joint project will be carried out in a situation in which

[10] I need not have mentioned my cousin previously in the conversation for him to be salient. Perhaps me and my conversation partner had just visited him, and he accomplished some feat which made his tallness salient. I will come back to this point below.

[11] There is a small industry dedicated to characterizing assertion. Two important recent monographs are McKinnon (2015) and Goldberg (2015). Cf. also the collections organized by Brown and Cappelen (2011) and by Goldberg (2020).

certain information is taken to be shared by the participants in the project, and will be guided by that body of information" (Stalnaker 2014, p. 6).

This allows us to re-think the simple communicative situation described by Lewis in our first section as a very impoverished instance of joint activities, in this case, a simple exchange of information. What the new framework shows is that most communication takes place against considerably richer backgrounds, in particular, most communication is embedded in a wider variety of goals. Moreover, it also allows us to think of coordination in terms of *joint projects* and *joint activities*, that is, activities in which the participants are not merely *alternating* autonomous actions (as in the simple encoding-decoding scheme proposed by Lewis in the first section above), but, rather, *collaborating* on the same action, an action that depends on everyone doing their part in order to be completed. This is important because, as Clark (1996) emphasizes, communication is *also* a type of joint activity in which *what is said* by the participants is *jointly construed* during the activity.

That is not to say that there is no specific linguistic element to communication; that would be absurd. Here, Lewis notion of *conversational scorekeeping*is particularly useful. The basic idea of conversational scorekeeping is essentially the same as Stalnaker's ideas summarized above: conversation is a rational, goal oriented activity and, as such, the participants in a conversation need to keep track of the shared information in order to better coordinate (or compete, since nothing prevents the conversation from being adversarial, as in a debate) among themselves. Like in any game, however, the evolution of the play is dictated not only by the body of information available to the participants, but also by the rules of the game, which specify the permissible moves given the current state of the play. That is, in additional to the rational constrains which are common to any joint endeavor, Lewis considers also the *specific* rules—i.e. the specific *linguistic conventions*—that inform the dynamics of conversations.

This new element may affect the mechanisms by which speakers interpret one another. Take, for example, how hearers determine to which person a speaker is referring to by the use of "She" in "Olívia worked a lot today. She must be tired." According to Stalnaker's account, the first sentence of that piece of discourse introduces Olívia as a topic in the conversation, thus raising her salience, which in turn makes it reasonable for the hearers to consider the speaker to be referring to Olívia in the second sentence. Stojnić (2021, Chap. 6), on the other hand, argues that it is part of the conventional linguistic rules (which account for the coherence of the discourse) that the anaphoric antecedent of "She" must be the person mentioned in subject position in the previous sentence.[12]

[12]Actually, her account is more complicated, since it includes a hierarchy of prominence according to which, e.g., nouns in the subject position take precedence over nouns in direct object position: cf. Stojnić (2021, Chap. 5) for details. This is important because part of her argument for the linguistic character of this

Be that as it may, my claim is that whether one follows Lewis or Stalnaker in drawing the line between purely pragmatic processes and linguistically driven mechanisms, the determination of truth conditions of utterances will often fall on the pragmatic side of the divide. Consider an example, adapted from Travis (1998/2008, p. 135). Suppose a person utters "This pen is blue", and suppose it is reasonably clear, perhaps by way of Stojnić's prominence logic, to which pen the person is referring to. In order to evaluate whether the person's claim is true, we need to know if that pen is indeed blue. But blue in what sense? Is it blue in the sense that its external color is blue? In the sense that its ink looks blue? In the sense that it writes in blue? If the latter, does it still count as blue if the ink disappears after a few seconds? There is now way of settling this question without knowing the *purpose* of classifying the pen as blue. If, say, the purpose of the classification is to organize the pens a person uses to highlight passages in a text according to different color schemes, it matters that it writes in blue (without disappearing). If, say, the purpose is to organize a person's pens according to an external chromatic scale so that it forms an aesthetically pleasing arrangement, then what matters is if the exterior of the pen is blue.

Of course, one example is hardly enough to settle a thesis. Indeed, as Collins (2020, p. 4) notes, one runs the risk here of "mere intuition mongering", as if a bunch of examples simply spoke by themselves. That is why Collins chooses to argue for the same thesis I am here defending by considering in detail syntactic descriptions of a range of sentences, showing how they fail to fully determine the content of what is said. While I definitely commend this "bottom up" strategy, here I want to propose a more abstract consideration. In the previous example, in order to arrive at the correct truth conditions for the utterance, one needed to consider it in the context of the larger purpose that the utterance served. My point is that this feature is not an accident.

As we have seen, both Stalnaker and Lewis emphasize that speech acts, as acts, are often embedded in larger action contexts, and are therefore goal oriented.[13] This means, first, that their correct evaluation depends on which goals they are serving, and which goal this is is a type of extra-linguistic information. Second, and more importantly, the goal may be jointly settled upon by the participants in the conversation, in such a way that both the goal and the means to achieving it *may be open to negotiation*. But, if the standard by which a given utterance counts as true depends on the goal of the utterance, and this goal may be open to negotiation, this means that what counts as the truth conditions of a given utterance

resolution of anaphoric reference is that it is a peculiarity of *English* that the hierarchy gives precedence to nouns in subject position. For my purposes, however, the above simplified example will suffice.

[13]Davies (2018) also stresses this fact, but uses it to slightly different purposes. Davies proposes that, given this goal-oriented structure of communication, expressions acquire "practico-normative properties" that allow hearers to interpret them as the speaker intends. I think this is certainly part of the story, but it is not all, since it ignores the role of grammar.

may also be open to negotiation. In the above case of the pen, for instance, perhaps which of the classifications schemes was to be employed became clear only *after* the utterance was made, through a negotiation process by way of which the participants *settled* on a goal.[14] Indeed, I would argue that this is precisely a case of Clark's (1996, chap. 7) *joint construal*, in which the participants of the conversation work together to settle on the meaning of an utterance. If the truth conditions of the utterances are open to negotiation, however, then they cannot be dictated solely by linguistic conventions. It follows that they must often fall on the pragmatic side of the divide.[15]

Interestingly, this points to a sort of inverse to the phenomenon we considered in the last section. There, we saw how Lewis's analysis of language in analogy with signal conventions failed because of the problem of how speakers interpret novel utterances. Here, we can see how it also fails even in the case of *old* utterances. Since the truth conditions of an utterance are open to negotiation in the sense discussed above, it follows that old utterances may have different truth conditions in different contexts—indeed, even in the same context, depending on how one construes that context. Words are, in Travis's terms, *plastic*:

> English furnishes descriptions that speak of given ways for things to be. The ways to be they speak of admit of understandings. So, in giving one of those descriptions, one might say any of many things, all the while using given English words as meaning what they do. That is one way plasticity allows for expressing novel thoughts. We come to see—the world may teach us—novel understandings of being thus and so, where our language provides us one or more descriptions which are, specifically, descriptions of being *that*. For any understanding of a thing, or things, being that, we may, in so describing things, say them to be as they are if that on that understanding. We learn new things one might understand by being that; new things being that might be taken to be. We thereby gain the ability to use old descriptions (ones already at our disposal) to new ends; for expressing new thoughts. (Travis 1998/2008, pp. 135-136)

One therefore sees how taking seriously Lewis's (and Stalnaker's) picture of communication as a rational, goal directed activity actually undermines his claim that meanings, in

[14] A similar view to mine is suggested by Heck (2014, pp. 340ff), though he does not quite put it in the same terms. Heck also emphasizes that speech acts must be located in larger action contexts and that this requires that, in order to successfully communicate, "there is work to be done on both sides" (Heck 2014, p. 341). But he does not emphasize that this means that there is work to be done on both sides *to settle upon the right truth conditions* for the utterance, though I believe this is implicit in what he says.

[15] I say "often" because there may be rare cases in which the truth conditions are solely determined by linguistic conventions. Collins (2020) himself takes some weather reports to qualify, though he also argues that the truth conditions so determined are in many ways deviant.

the form of truth conditions, are determined by linguistic conventions. Ludovician conventions are much too rigid to account for the plasticity of our words. That is not to say that "everything is permitted", though. In the next section, I will sketch an account of how grammar constrains what is said by a given utterance by imposing a *structure* on what is said.

2.5.2 Grammatical Structure

In the last section, I argued that truth conditions are not linguistically determined, in the sense that many pragmatic factors may enter into their determination. Hence, insofar as meanings—at least as objects of semantics—are, by hypothesis, linguistically determined, meanings cannot be truth conditions. But what are meanings, then? And how do they constrain the interpretation of a sentence?

Before answering those questions, let us go back to the picture outlined in the last section. There, following Lewis and Stalnaker, I argued that communication is a kind of collective goal directed rational activity. This meant that what counted as successful communication was in part determined by those participating in the conversation, be it by their implicit behavior (e.g. by continuing the conversation) or else explicitly, as when misunderstandings arise. This has two consequences for how we conceive of meanings.

It is a common assumption that successful communication depends on the hearer grasping the exact content of the speaker's utterance, as if this content were always available before the communicative process takes place.[16] If the reasoning so far is correct, this assumption should be rejected: participants in a conversation do not necessarily come equipped with a previous understanding of sentences when entering a communicative situation, and, if they do come so equipped, it is not important that such an understanding be shared, as they may modify and build together a common understanding *during* the exchange. Moreover, the standards by which we assess whether such a common understanding depends on the goal of the activity in which communication is embedded, so that what seemed to be a common understanding in one particular occasion may need to be reassessed in a different occasion.[17]

Second, it also means that language is a *tool* that can be flexibly adapted to a variety of purposes. In many cases, a speaker means not to settle on a *single* interpretation for her utterance, but, rather, offers the utterance as an *invitation* for further exploration—this is especially salient with literary language, though, of course, not restricted to such language.[18] If that is so, however, then meanings cannot be such as to rigidly determine one fixed

[16] Cappelen and Lepore (2005, chap. 8) press exactly this point against the "radical contextualist", i.e. that she cannot account for the fact that it is the same proposition that is shared by speaker and hearer.

[17] This is connected to Travis's (2017) point that whether a thought counts as the same or not depends on the occasion in which the assessment is being made.

[18] This point is emphasized by Lepore and Stone (2015b, esp. Part III).

interpretation for an utterance. On the contrary, there must be, to use Travis's term from the last section, a certain *plasticity* inherent in them, leaving room for open-ended exploration.

One important piece of this puzzle was provided by Pietroski (2018). It is a common suggestion in the Chomskyan tradition to treat meanings as being intimately related to concepts, for instance by treating them as instructions to fetch or build concepts.[19] Pietroski adds to this idea that the instruction may not be to fetch a *specific* concept, but rather *a* concept from a *family* of concepts. Moreover, such concepts *do not* come with a syntactic type attached, so that a word such as "dog" can serve as an instruction for fetching a concept of an animal or of an action, depending on the syntactic environment in which it appears.[20] Words, then, do not necessarily come with a grammatical category attached, but *assume* a grammatical category depending on the position in which it is inserted in a sentence. This explains at least in part the plasticity of language: since words are not attached to a *single* concept, but to a *family* of a concepts, different understandings (even by the same person) may correspond to different concepts being fetched each time.[21]

As I mentioned, this initial polysemy is partially constrained by the syntactic environment in which the word appears. As the word assumes a grammatical category depending on its position in a sentence, conversely, the position of words in a sentence *constrains* the interpretation we may give it. So, for example, depending on its position in a sentence, the same word may be interpreted as the agent or patient of an action, or even as the action itself. In the case of "dog", if it appears in subject position in an active voice sentence, such as in "A dog chased a man", then the word must be taken to mean the agent of the event, thus suggesting that it is the animal which is in play. On the other hand, if it appears in the verb position, such as in "The paparazzi dogged her mercilessly", then it must be interpreted as a verb. Indeed, in this last sentence, we may also see that the sentence frame forces one syntactic item ("The paparazzi") to be the agent of the construction, another ("her") to be the patient of the construction, and still another ("mercilessly") to denote the way the action was done. So this particular frame instructs us to build the concept of an action that was done by someone on something in some manner. Interestingly, what is at first sight a constraint also adds flexibility to the language, since inserting a given word into a syntactic position may *coerce* it into assuming a syntactic feature that is not usually associated with the word, as in, e.g., "I windowed the north wall" (Borer 2005a, p. 8), thus allowing us to build

[19] Cf. McGilvray (1998) for an early defense of this idea and Boeckx (2010, Chap. 7) for a survey, including references.

[20] Here, Pietroski is following Borer (2003, 2005a,b, 2013) and her exo-skeletal approach to the lexicon. I will come back to this below.

[21] This *partially* explains the plasticity, since another point is that even the same concept may admit of different understandings each time it is employed.

novel concepts on the fly.[22]

In any case, what this means is that the syntactic structure of the sentence *imposes* a certain conceptual organization on what is said, by, for instance, attributing the role of agent to a certain concept, of patient to another, etc. That is why authors such as Boeckx and Pietroski do not merely talk of *fetching* concepts, but also of *building* them. The conceptual organization imposed on what is said may be wholly novel, so that the syntactic structure of the sentence is more akin to a recipe for building a concept than a recipe for merely fetching concepts. That is why lexicalization is not just the expression of thought, but also its organization.[23]

The picture which emerges is this. In uttering a sentence, a person is *inviting* a person to entertain some thought. This thought is conceptually organized by the syntax of the sentence in question, being given, for example, an event structure with agent and patients. This structure, in its turn, is filled in by concepts indicated by the words used, though *which* concept from a given family fulfills this role is constructed by the hearer, obeying the strictures of the recipe indicated by the syntactic structure of the sentence. A person's internalized grammar is, then, precisely this capacity for conceptual organization following syntactic lines, in such a way that the *stable* part of meaning is wholly internal to grammar.

Notice that this lines up well with the open ended conception of communication sketched above. In a conversation, the speaker is not forcing a certain understanding on the hearer, but rather inviting her to build this understanding on the basis of her internalized grammar. This is very similar to the language acquisition strategy known as *syntactic bootstrapping*, in which the person acquiring a certain piece of vocabulary uses the grammatical environment of the piece to infer its meaning.[24] Indeed, one way of putting my point is to say that syntactic bootstrapping is a *pervasive* strategy, not confined just to the early stages of language acquisition.

2.6 Conclusion

We started with Lewis's code model for language, according to which linguistic communication is essentially a matter of encoding and decoding sentences in order to retrieve pieces of information. This set up a tension inside Lewis's account of language, since, on the

[22] I am obviously oversimplifying a lot. Cf. Borer (2003) for an introduction to this phenomenon and to her exo-skeletal approach.

[23] Indeed, theoreticians such as Reboul (2017) will argue that *this* is the primary function of language, even in terms of evolutionary history: to organize thought.

[24] Cf. Landau and Gleitman (1985, Chap. 7) for a description of this process in relation to a blind child's acquisition of verbs related to sight, and the essays in Gleitman (2020, Part IV and V) for the theoretical framework and more empirical evidence.

one hand, the code model suggested an account, officially adopted by Lewis, according to which the *process* of encoding and decoding pieces of information was largely irrelevant to the nature of language, yet, on the other hand, in the face of many obstacles, Lewis needed to appeal to precisely such processes, in the form of grammars, to give a coherent account of language use. Moreover, this appeal to grammars sat uneasily together with his commitment to common sense psychology, putting more strain into his account.

An important part of the solution to this problem involved a strict separation between semantics and pragmatics. Common sense psychology, or something close to it, may help us to understand how people arrive at *truth conditions*, but truth conditions are an unstable, pragmatic phenomenon open to negotiation. Once we separate questions of meaning from questions of interpretation, we are better able to accommodate the open-ended character of interpretation and the stable character of meaning, while at the same time showing how meanings can be flexible enough to underwrite interpretation. One consequence of this is that semantics, as our theory of meaning, becomes decidedly parochial: specific facts about the human psychology enter both into the definition of what meanings are and into our account of language use. Perhaps, then, this may explain the fundamental tension in Lewis's project, namely his desire to explain what is a *parochial* dimension of our being as if it were an objective feature of *any rational being*. Language is not the house of being, nor even the house of rationality. It is enough that it be a good enough tool for *our* being.

Chapter 3

Davidson Between Convention and Interpretation

3.1 Introduction

Donald Davidson is justly famous for his bold proposal: a theory of meaning for a natural language should take the form of a theory of truth. Yet, there is in this proposal a certain ambiguity. Is the proposal that a theory of truth should specify the *systematic and stable* semantic features of the language? Or is the proposal that a theory of truth should specify the *systematic ability* of interpreters to attribute truth conditions to others' utterances? For reasons that will become apparent, let us call the first proposal the *Convention Proposal* and the second the *Interpretation Proposal*.

Of course, one might argue here that interpreters have a systematic ability to attribute truth conditions to others' utterances *because* they exploit the systematic and stable semantic features of the language. The main thesis of this article is that this road is closed off to Davidson, and that he thus faces a dilemma. As I will argue below, Davidson himself gives us the tools to recognize that, insofar as his theory aims at describing truth conditions, it will not be about the stable semantic features of the language, since, as he later argues, truth conditions are the result of interpretation, and are therefore not stable. So *either* his theory is a theory about the stable semantic features of the language, but then they are not about truth conditions, *or else* they are about truth conditions, but then they are not systematic, and therefore cannot assume the form an axiomatic truth theory. Davidson is aware of the dilemma, and he embraces the second horn (which is effectively the Interpretation Proposal), that is, he agrees that his theory is not about the stable semantic features of the language. He nevertheless attempts to deflate the consequences of this move by arguing that an axiomatic truth theory still describes the theory the hearer uses to interpret the speaker, but this is too much of a strain—or so I shall argue.

Now, from a historical point of view, it is interesting to note that Davidson actually started by defending a version of the *Convention Proposal*. Although this is not very explicit in his early published work,[1] it is clear in the recently published 1970 John Locke lectures that Davidson thought of a semantic theory for a language as aiming at capturing its stable and *conventional* semantic features, or, as he puts it, to describe its "*conventional semantic core*" (Davidson 1970/2020, p. 26, his italics). It is only later, by the time of his debate with John Foster,[2] that Davidson eschewed any appeal to conventions when theorizing about language: "just as *Lear* gains power through the absence of Cordelia, I think treatments of language prosper when they avoid uncritical evocation of the concepts of convention, linguistic rule, linguistic practice, or language games" (Davidson 1976/2001, p. 171). It is also by that time that we find, in the place of convention, an appeal to *interpretation* in setting the aims of the theory: a theory of meaning for a language must "explicitly [state] something knowledge of which would suffice for interpreting utterances of speakers of the language to which it applies" (ibid.). In other words, by 1974, Davidson had switched to the *Interpretation Proposal*.

What happened between 1970 and 1974? Ian Rumfitt (2021), in his recent review of the John Locke Lectures, suggested an answer: "Radical Interpretation" (Davidson 1973/2001b) happened. In that essay, Davidson broadened his scope by considering how a theory of meaning for a language could be integrated into a unified account of a person's behavior—i.e. into a global *interpretation* of the person—, thus bringing together this project in line with his other major project at the time, namely to develop an account of intentional action. But, as we will see, if the aim is to interpret a person's behavior—to place it into a larger network which exemplifies a *constitutive ideal of rationality*—, then, since much of this behavior is *novel* and *idiosyncratic*, appeals to conventions cannot be explanatory, since those appeals presuppose a regularity that is simply not there.[3]

Unfortunately, this seems to undercut the rationale behind Davidson's semantic theory. The first theory had it right: a semantic theory, if it aims to be systematic, should be a theory about the stable semantic features of the language. It is impossible to give a *systematic* theory about how to *interpret* other people—a point that Davidson himself repeatedly makes. Unsurprisingly, by the time he shifted to the Interpretation Proposal, he also dropped the idea that a semantic theory should aim at capturing the explicit knowledge of language users, and started adopting the rather odd formulation that the theory proposed should only be *sufficient* for interpreting another, even if it is never actually employed. It is and odd formu-

[1] But do see Davidson (1969/2001, pp. 44-45) for hints of this view.

[2] This debate took place in 1974 but was only published in 1976.

[3] This point, of course, was already emphasized by Chomsky (1959) in his famous review of Skinner. Davidson (1984/2001, pp. 277-278) deploys it in the context of his criticism of Lewis's (1969, 1975/1983) theory of language as a type of conventional signaling system.

lation because, if the goal is to place a person's linguistic utterances in a pattern constituted by an ideal of rationality, then it seems we are interested in the principles and competences *actually* employed by the person, not in a theory completely divorced from linguistic reality. Indeed, in "A Nice Derangement of Epitaphs", Davidson himself recognizes this. There is then a strange wavering on Davidson's part as to whether or not the theories he describes should be attributed to speakers or not.[4]

In light of this, Rumfitt, in the review mentioned above, suggests that we go back to the first Davidson, that is, that we adopt the Convention Proposal. This, however, is to ignore a central insight of the later Davidson: there is a connection between *interpreting* another and attributing truth conditions to this person's utterances, and this connection *cannot* be explained in terms of conventions. In other words, whereas there is a connection between interpreting another and truth conditions, there is no such connection between conventions and truth conditions. The solution I propose is to follow this insight to its limit: this means that there can be no *systematic* theory of truth conditions. The first Davidson was right that a systematic semantic theory must be about the *stable core* of language, but wrong in associating this core to conventions and truth conditions. The second Davidson was right that there is a connection between truth conditions and interpreting another person, but wrong in supposing that this could be the object of a systematic truth theory.

This essay is thus structured as follows. In the first part, I will analyze Davidson's first proposal, especially as presented in his John Locke Lectures. Here, the emphasis is on his idea that a truth theory could capture what he then called the *conventional semantic core* of a language. Afterward, I will analyze his theory of interpretation as developed in the famous essay "A Nice Derangement of Epitaphs", together with his concomitant rejection of any role for conventions in analyzing a language. Finally, I will compare the new theory with the old theory along a number of evaluative axes, noting that, although the new theory is correct as far as it goes, only by dropping the assumption that it is dealing with some kind of *formal semantics* we arrive at a theory that is a marked improvement over the old one. This allows me to draw out what I take to be the main lesson from this discussion, namely,

[4]Lepore and Ludwig (2005, Chap. 9) also note this ambivalence, yet they defend that Davidson never abandoned the idea that a theory of truth should correspond to a person's linguistic abilities, effectively adopting the line proposed by Davies (1981) and Evans (1981/1985), according to which each axiom of the theory should correspond to a disposition of the speaker. As I explain below, the situation is more complex. As I argue in sections 2 and3, the first Davidson (up to 1970) does consider that a theory of truth should describe a person's linguistic abilities, but these abilities are not a form of "tacit knowledge", but rather *explicit* knowledge of linguistic conventions. For the reasons I discuss in section 4, *late* Davidson (after 1974) eventually abandoned the idea of linguistic convention, thus arriving at a position closer to the one described by Lepore and Ludwig, according to which the theory describes the linguistic dispositions of the speaker, but the situation is made complicated by the fact he *also* wants the theory to be *used* by the interpreter to attribute truth conditions to the speaker's utterances.

that truth conditions are a *pragmatic* phenomenon and that, insofar as semantics aspires to be a systematic science, it cannot treat of truth conditions.

3.2 The Learnability Requirement and the Systematic Character of Linguistic Competence

From the beginning, Davidson defended that a finitely stated theory of truth for a natural language was the best way to meet a constraint on theories of meaning he had already proposed in an early paper (Davidson 1965/2001), namely that such theories should be *compositional*, that is, they should state how the meanings of complex expressions systematically depend on the meaning of their parts.[5] The central argument for this constraint is based on the *Learnability Requirement*: a theory of meaning for a language should not imply that the language is unlearnable. Since Davidson's reasoning is revealing of a tension his position at the time viz. what theories of meaning are theories of, a tension that will be resolved in one way in the John Locke Lectures and in another way in "A Nice Derangement of Epitaphs", it will be worth to unpack this idea.

Davidson starts by noting the fact that, if one accepts that a theory of truth in the style of Tarski can do duty for a theory of meaning, then we have one way of specifying the meaning of a sentence as a function of its parts. The idea is well-known: since the recursive clauses of the theory specifies the truth conditions of complex sentences in terms of the truths of their parts, one could see how the meaning of a given complex sentence depends on the meaning of its parts by simply following the derivation of the sentence's truth conditions from the axioms of the theory.[6] Now, notice that, so far, the discussion has been rather abstract. A theory could meet this requirement even if it were non-finitely axiomatizable. The only requirement here is that the theory must specify the meaning of complex expressions on the basis of their parts. In contrast, Davidson's next movement shows that he is clearly not interested in such abstract theories.

After briefly remarking on the idea that a theory of truth could serve as a theory of meaning, Davidson then mentions that this theory could help explain how "an infinite aptitude can be encompassed by finite accomplishments" (Davidson 1965/2001, p. 8). An "infinite aptitude": the capacity to speak a natural language is *productive*, i.e. there is no

[5] There are some subtleties related to how to best formulate the compositionality constraint, but they will not matter here. Cf. Szabó (2000) for details.

[6] There are some complications here, since it is not initially clear when the conclusion of a proof in the theory actually gives the truth conditions of a sentence or simply a "garbage" theorem. In order to filter these "garbage" theorems out, those who followed Davidson introduced the idea of a *canonical proof*, limiting the logical rules which could be applied in a proof. Cf. Lepore and Ludwig (2005, Chap. 7) for discussion.

upper bound to the number of sentences from a natural language a person is capable of understanding. "Finite accomplishments": if we restrict our theories to *finitely axiomatized* theories, then it would suffice to learn finitely many axioms to understand the whole theory. If, therefore, we hypothesize that our theory of meaning is at the same time a theory of a speaker's *linguistic competence*, we would be able to explain how speakers are able to understand an infinite number of sentences. As a bonus, this very restriction forces us to impose considerable structure on the theory, so that we can, at the same time, also explain the language's *systematicity*, i.e., the fact, roughly, that if a person can understand a sentence with a given number of primitives, then she can also understand other sentences built using these primitives.

One important distinction to be made at this point is between productivity and systematicity, on the one hand, and creativity, or the ability to understand novel sentences in different contexts, on the other hand. What has been said so far is compatible with a person learning a new language *ab initio* in every new context by learning new primitives in each such context. Of course, if the competence described by our theory is *stable*, that is, if it can be brought to bear in novel contexts, then the fact that our competence is systematic is also able to explain how we handle novel sentences, namely by noting how the novel sentence exploits the primitives we already know. But nothing Davidson has said so far points in the direction of these capacities being stable, so that the door is apparently opened for each new context requiring a different theory. This possibility may be initially counter-intuitive, but it is exactly the position adopted by Davidson in his later work, "A Nice Derangement of Epitaphs".

Now, in the John Locke Lectures, as we will see, Davidson adopted a different position, according to which our linguistic capacities *are* stable. Some indication that the same view is operative in this essay can be found in Davidson's appeal to *rules* when specifying the type of linguistic competence he is aiming to capture. When discussing unlearnable languages, for instance, he says that such languages contain sentences "whose meanings are not given by the rules already mastered" (Davidson 1965/2001, p. 8). Similarly, when discussing his empirical assumptions, he lists among them the assumption that we do not have the capacity to "intuit the meanings of sentences on no rule at all" (Davidson 1965/2001, p. 9). This is far from decisive, since to say that our linguistic competence consists in the mastery of rules is not necessarily to say that such rules are stable across contexts, much less that they are, for example, conventional.

The problem is how to interpret the notion of a *rule* in this context.[7] If Davidson means by "rule" simply an axiom in a formal system, he is clearly not committed to their being

[7]Jackendoff (2003, p. 55) provides a nice catalogue of the options in the context of discussing the nature of grammatical rules in the generativist project.

diachronically stable. If he means something thicker, like a norm or law which governs a patterned activity, then they must be diachronically stable and probably conventional. Clearly, how we interpret this notion also has consequences for how to understand Davidson's linguistic enterprise. The first interpretation is much closer to the deflated understanding Davidson will put forward after 1974, whereas the second interpretation is closer to the idea that the theory devised by the semanticist must be (implicitly or explicitly) known by the speaker, a position espoused by Davidson in the John Locke Lectures. Although the evidence is flimsy, it seems to me that Davidson's talk of *mastering* a rule points more to the second interpretation, as, typically, what one *masters* is a skill, an ability that can be redeployed in different contexts. If so, then, in this early article, Davidson would be drawing a tighter connection between systematicity and creativity than in his later articles.

Regardless of how Davidson understood the term in 1965, there is one important point in which these remarks conflict with his later position. Davidson is assuming here that we *always* intuit the meaning of sentences by following a rule, be this rule just a formal axiom, a conventional norm governing an activity, or something else. As remarked by Gustafsson (1998, p. 448), this assumption is rejected in "A Nice Derangement of Epitaphs", since there Davidson is quite explicit that we *can*, after all, understand the meaning of sentences without recourse to rules. This will introduce the central tension I will explore in this article, but, before getting into that, there is still some legwork to do.

In particular, there is one further assumption from this article that I would like to comment on, namely, that our capacity to understand a language *is* productive in the right way. It could very well be that *languages* are productive, in the sense of allowing for an infinite number of sentences, yet our capacity to understand the sentences of such languages be bounded. In those cases, we would only ever learn *part* of a language, and not the whole language. Davidson, however, had already stated in the end of the first section of his article that this is not a coherent possibility:

> The lesson for theories of language learning is wholly negative, but not perhaps without importance: in so far as we take the 'organic' character of language seriously, we cannot accurately describe the first steps towards its conquest as learning part of the language; rather it is a matter of partly learning. (Davidson 1965/2001, p. 7)

Admittedly, Davidson's reasoning is a bit obscure here. Nevertheless, the point seems to be that languages are 'organic'[8] in the sense that the meaning of a given sentence is systematically connected to the meaning of other sentences in the language. Thus, to learn the meaning of a single sentence one must, *ipso facto*, learn the meaning of many other

[8]The expression is Strawson's. Cf. Strawson (1956, p. 452), quoted by Davidson (1965/2001, p. 7).

sentences in the language. Of course, as Lepore and Ludwig (2005, p. 30n16) note, this only allows us to conclude that, at best, we cannot learn sentences in isolation, but in chunks. Nothing so far allows us to conclude that there is only *one* chunk, that is, the whole language.

If we assume systematicity, on the other hand, it does seem that we have at least a good heuristic to conclude that there is only one chunk, after all. The reason is that, in order to finitely specify the axioms correctly, one must consider the effect of a single axiom in *all* the sentences in which it appears. Since the systematicity property of a language implies that any grammatical combination of the primitives has to be accounted for, this means that the meaning of any single sentence may indirectly constrain the meaning of all the semantic primitives of the language, as any primitive not in the sentence could potentially occur with a primitive that *is* in the sentence, or with a primitive that occurs with a primitive that is in the sentence, etc.[9] The ripple effect may be that the meaning of any sentence, in virtue of the constraint imposed on the meaning of the primitives, is connected to the meaning of every other sentence in the language, so that the whole language is one big organic chunk, as required. This is not a *proof* of this conclusion, but it is suggestive of the reasoning behind Davidson's assertion.

In any case, this is important because it imposes a very serious constraint in linguistic theorizing, i.e. that a theory of meaning cannot make recourse to the idea of someone learning just part of a language. Consider, for example, dialects. One could consider a dialect as a part of the main language, perhaps with some local differences, but this is ruled out by Davidson's holism, since there are no local differences to him—every difference is global. Similarly, if you and I have different vocabularies, so that out linguistic competence is described by different primitives, then we have different languages—each with out particular idiolects, perhaps. Indeed, if different time slices of the same person has a different vocabulary, again, these different time slices would have different idiolects, in such a way that talk of a *person's* idiolect seems to be mistaken. If that is so, however, in what way can we talk about *languages* at all? What this reveals is that Davidson's holism already pushes in the direction of denying that public languages, or even idolects, as a diachronically stable units, are explanatory relevant at all.[10] Obviously, at this point, this was not explicitly on the table yet—indeed, Davidson's talk of "partly learning" a language (but what is to partly

[9]Camp (2016) presents a similar reasoning: "Thus, apparently trivial, possibly homophonic T-sentences gain theoretical traction not in isolation but as a whole, as theorems to be generated by an adequate compositional semantic theory. More importantly for our purposes, this means that the meaning of an individual word is constituted by its contribution to the truth-conditions specified by all of the T-sentences whose quoted sentences contain it." (Camp 2016, p. 119).

[10]The point that even the relevance of idiolects is called into question is important, since some readers took the main thrust of "A Nice Derangement of Epitaphs" to be to argue *in favor* of idiolects, as opposed to public languages. Cf. Begby (2016, p. 46n39) on this point.

learn if not to learn a part?) seems to blunt the force of this point considerably—, but the tendency was already present.

It is therefore clear that the systematic character of language—the fact that repeatable features have the same semantic contribution in each sentence that they appear—is central to Davidson's account.[11] Still, as we have seen, it is not entirely clear the scope of this systematic character. On the one hand, Davidson's talk of mastery of rules seems to point in the direction of diachronically stable conventions, thus also covering uses of sentences in novel contexts. On the other hand, the same systematicity, unfolded into his radical holism, seems to point in the direction of abolishing talk of diachronically stable units altogether. In the next section, I will show how the John Locke Lectures embrace the first option, while, in later sections, I will explore the second option.

3.3 The Conventional Semantic Core of a Language: Davidson's Proposal in the 1970 John Locke Lectures

The last section explored a tension between two demands that seem to fall out of the systematicity requirement espoused by Davidson, demands that pulled in opposite directions when it came to the diachronic stability of the semantic rules employed by speakers of a language. There, I also mentioned how, in the John Locke Lectures, Davidson espoused the view that systematicity was connected to creativity, and, therefore, that the rules formulated by the semantician should be considered as diachronically stable. Thus, we find him claiming that:

> It is popular nowadays to stress the creative aspects of language, the fact that someone who can speak or decipher a language can cope with sentences he has never heard. This is usually taken, and I think rightly, to imply that to grasp a language is to deal with it as a recursive structure. (Davidson 1970/2020, p. 29)

Obviously, one could take Davidson to be simply staking out a possible material implication between the claim that speakers are creative and the recursive[12] structure of the language, but it seems to me fair to read the above quotation as claiming, first, that speakers *are*, in fact, creative, and, second, that the *explanation* for creativity is the recursive structure of their languages. Although he does not develop the point, it is reasonably clear how the

[11] And rightly so, I might add. Gustafsson (1998) believes that later Davidson should jettison systematicity completely in light of his arguments against conventions, yet this is to throw the baby out with the bath water. Rather, what one should give up is the idea that repeatable features contribute to the determination of *truth conditions*. I will have more to say on this later.

[12] Incidentally, Dummett (1986, p. 459) is obviously right when he points out, against Hacking (1986, pp. 455ff), that, by "recursive", Davidson simply means "specified inductively".

explanation would go, *granted the assumption that the rules of the theory are stable*. Suppose that our grasp of a language is recursively specified, and that the linguistic competence specified by the axioms of the theory are stable across contexts. Then, upon encountering a sentence she has never heard, a subject would be able to piece together its meaning from her knowledge of the meaning of its parts and of its mode of combination. Indeed, if her semantic competence were to be modeled by an axiomatic truth theory for the language, the subject's (perhaps unconscious) reasoning could also be modeled by a canonical derivation of a T-sentence from the axioms.

I will come back to the requirement that the axioms of a theory describe diachronically stable competences in the next section, when considering Davidson's argument *against* that position. For now, I will focus on *why* these competences are stable. The reason is simple: the competence to speak a given language has as its core knowledge of linguistic conventions. As Davidson puts it, this *conventional semantic core* of "a particular language *is* what the theory [of meaning] describes" (Davidson 1970/2020, p. 27). Given, then, that our semantic competence consists in knowledge of linguistic conventions, we have a ready explanation for the diachronic stability of this competence, as conventions, on any reasonable account, are *regular* patterns of behavior—the dispute is about what *else* conventions are.[13] Fortunately, such disputes need not detain us here, as it will be precisely this regularity aspect which will interest us in Davidson's account.

Given, then, that conventions are (at minimum) regular patterns of behavior, one may wonder in what kind of behavior these linguistic conventions consist. Since they are *linguistic* conventions, they must be conventions about speech acts. Since they are *conventions*, they must be about *regular* speech acts, acts that are regulated by certain rules which can be *systematically* codified. Davidson believes only one kind of speech act satisfies that requirement, namely *speaking the truth*. The reason he gives for this is very revealing and worth pausing over for a moment:

> It is hard to think of a single further and interesting case [besides speaking the truth] of a linguistic action which is such that, simply by knowing the language, one can state a criterion determining whether that action has been performed. Asserting, ordering, insulting, promising, describing, warning, instructing, questioning, and stating are actions that can be performed by those who have the gift of speech, but in not one of these cases, I think, can knowledge of the criteria that determine whether one of these actions has been performed be called exclusively linguistic knowledge. (Davidson 1970/2020, p. 21)

[13] Cf. Rescorla (2019) for various accounts of conventions. It is clear from Davidson (1984/2001, pp. 276ff) that, whatever misgivings he may have about Lewis's account of conventions, he accepts that they are regularities of behavior.

There are two different claims here:

1. Every speech act, with the exception of speaking the truth, requires more than linguistic knowledge in order to determine whether it has been performed.

2. In order to determine whether someone spoke the truth, it is possible to rely solely on linguistic knowledge.

Let us tackle each of those claims in turn, starting with the point that we need more than linguistic knowledge to determine whether a speech act other than speaking the truth has been made. Since Davidson's argument is rather opaque (to me at least), I will pursue an indirect strategy of introducing some extraneous machinery in order to clarify what is, in my view, the main thrust of his position. Hopefully that will not result in too much violence to the text.

In considering speech acts, it will be useful to have a tractable representation of the *context* in which they are made.[14] Without going into formal details, I propose that we model such a context with an abstract representation having at least three main components:[15]

1. For each participant in the conversation, the context will include a *deontic score*, that is, a representation of each participant's *commitments* and *entitlements*.

2. For each participant in the conversation, the context will include a *mental score*, that is, a representation of each participant's mental states (which we can simplify to belief-desire psychology).

3. The context also includes the *common ground*, that is, the pool of information that is mutually expected by each participant to be shared with all others in that conversation.

Using this idea, we can then think of a conversation as evolving along the following path. Suppose I promise Olívia to go to the park with her on Wednesday. This has an effect on both my deontic score and on hers: I become committed to going to the park with her on Wednesday, and she becomes entitled to hold me to this commitment. If I am truthful and she trusts me, this also changes our respective mental states, say, by making us both

[14]There is an enormous work on speech act theory, of which I am going to merely touch the borders. For a very nice overview, cf. Harris et al. (2018), and, for an overview centered on assertion, Sbisà (2020).

[15]I am here adapting a proposal by Antonsen (2018) to unify two distinct traditions in speech act theory, the *commitment* account, chiefly developed by Brandom (1994)— cf. Shapiro (2020) for an overview—, and the "change-to-the-common-ground" one, chiefly developed by Stalnaker (1970/1999, 1974/1999, 1978/1999, 1998/1999)—cf. Clapp (2020) for an overview. The adaptation is to include each individual's mental states—as distinct from their mutual expectations as represented in the common ground—, in order to make room for Harris's (2020b) position that often the point of a speech act is to change the *addressee's* mind, not to update the common ground.

desire to go to the park on Wednesday and believe that we will do so. Finally, if there is enough trust between us, it will also add the preceding information to the common ground, so that each of us can mutually expect that, for instance, I am committed to going to the park on Wednesday with her, that we each believe that I will do so, etc.

Following Austin (1962, pp. 120ff), we may distinguish two dimension of speech acts.[16] First, there is the *illocutionary force* of the act, which I will here define in terms of the change in the commitments and entitlements of the context.[17] Second, there are the *perlocutionary effects*, which we may think as the changes to each participant's mental states and to the common ground brought about by the act. It is thus possible to individuate speech acts by focusing either on its illocutionary force, on its perlocutionary effects, or both. An assertion, for instance, could perhaps be defined as an act in which I aim to make others believe in its truth (a perlocutionary effect) and in which I also undergo a commitment to its truth (an illocutionary effect).[18]

Obviously, the act need not be successful in order to be the act it is, but it is nevertheless defined by its success conditions, that is, by the effects it is intended to have.[19] Still, there is one kind of failure that *does* result in the act not being performed at all, namely failure of *uptake*.[20] For our purposes, we can define uptake as the hearer's updated belief that the change in the deontic score proposed by the illocutionary force of the act has in fact taken place.[21] The reason why uptake is necessary for the act to be performed is that deontic scores own their very existence to the practical attitudes of those affected by them.[22] To be committed

[16] I will ignore the locutionary dimension for now.

[17] Austin actually refers to the "conventional effects" of the act. I think Sbisà (2002, 2020, §3.3) is correct, though, that the reason behind Austin's appeal to conventions was not to define an illocutionary act in terms of some kind of social regularity, but rather in terms of changes in the deontic score of the participants. In this sense, Harris et al's (2018, p. 2) complaint that Austin focuses on "highly ritualized examples" is completely besides the point. Basically, it misses the fact that Austin is not emphasizing the (supposed) conventional *means* of doing something, but rather the conventional *effects* of doing something. There need be no conventional means, much less a highly ritualized one, of bringing about a certain effect.

[18] This is just an example, and not intended as a real definition of assertion.

[19] As Mark Lance and Quill Kukla (writing as Rebecca Kukla) put it: "Crucially, then, the *output of a speech act is the normative statuses the speech act strives, as part of its function, to bring about*—not what it actually manages to bring about" (Lance and Kukla 2009, p. 16).

[20] Cf. Sbisà (2009) for a discussion of this notion from an Austinian perspective. There is some controversy on whether uptake is really necessary for the performance of a speech act—cf. McDonald (2021b, p. 3507-3508) and the references therein for a quick overview. For the reasons stated in the text, I agree with Hornsby and Langton (1998, p. 25) and McDonald (2021a, §II.1) that the idea of an illocutionary act as part of a *communicative exchange* points strongly toward the necessity of uptake.

[21] Geurts (2019, p. 18) gives a similar definition, but purely in terms of commitments: the uptake is defined to be the hearer's commitment to the fact that the speaker has committed herself to whatever is the output of the act. It seems simpler to define uptake simply in terms of beliefs, though.

[22] Cf. Brandom (1994, pp. 161) for discussion.

to something *just is* to have that status attributed to one by oneself and others. If the others do not recognize my commitment, then, *ipso facto*, there is no such commitment.[23]

The above analysis puts us in a position to state Davidson's two theses more precisely. They basically consist in the assertion that speaking the truth is unique in that it is not essentially defined either by its illocutionary force or by its perlocutionary effects. In other words, Davidson is saying that every speech act, with the exception of speaking the truth, must be defined by making reference to either its intended illocutionary force, to its intended perlocutionary effects, or both (the first thesis), whereas speaking the truth can be defined without making reference to either (the second thesis).

Supposing the preceding analysis is correct, we can better understand Davidson's argument for the special place of speaking the truth in linguistic studies. For suppose the two theses are true. Since every speech act other than speaking the truth is defined by its intended illouctonary or perlocutionary effects, it follows that, if an agent is in a context which somehow prevents those effects from taking place, the context will also prevent the act from taking place. As a consequence, in such a context, *no sentence which I utter will be sufficient for me to perform the act*. Hence, if there are such contexts, there will be no *purely linguistic* device that will guarantee, by its employment, that I will have performed the speech act in question. In other words, if there are such contexts, then there will be no purely linguistic conventions regulating the speech act.

Now, Davidson believes that there is at least one type of context that will *always* negate the force and the effects of my speech acts, namely pretending, in cases in which is part of the common ground that I am pretending. In those contexts, I can utter *any* sentence without thereby performing the speech act that I would perform in normal contexts, simply because, in those contexts, nobody will uptake my utterance as being of the speech in question, since they will believe that I am merely *pretending* to perform the speech act.[24] In other words, in pretense contexts, even if I intend to perform a given speech act, there will be a failure of *uptake* on the part of the audience, and therefore the speech act will fail to be performed.[25]

This seems reasonable, so that the first thesis is established. What about the second thesis?

[23] There are subtle questions here about whether uptake, and whether it is common ground that the uptake took place, should also enter into the success conditions for an illocutionary act. Cf. McDonald (2021a) for a very perceptive discussion.

[24] As Davidson notes, this need not be because I do not have the intention to perform the speech act. The problem is that I may be in a defective common ground, in which I believe not to be pretending anymore, whereas the others believe I am still pretending.

[25] Green (1997) argues that Davidson failed to establish his thesis, since, in pretense contexts, no speech act has been made, and the conventionality of linguistic devices applies only in cases where *some* speech act has taken place. I am unconvinced, for two reasons. First, Davidson's point can be precisely that no linguistic device guarantees that a speech act (*tout court*) takes place, something that Green does not dispute. Second, it is not clear to me that in pretense contexts *no* speech act has been made—say, narrating or joking.

Davidson's point is that it is not part of the definition of speaking the truth that it have any illocutionary force or perlocutionary effect. Consider the case of pretense. I can be part of a pretense and still, inadvertently, speak the truth (perhaps because I didn't know better). For example, I may be part of a stage play and say "Davidson was an American philosopher" as a line. Since this was part of the play, I do not undergo any commitment to the truth of this utterance, and my audience may not believe this utterance was true (perhaps they are all under the delusion that Davidson was British). Nevertheless, I have still spoken truly. Again, this is because whether or not I have spoken truly, according to Davidson, is not determined by either the illocutionary force or the perlocutionary effects of my utterance; for him, there need be no changes in either my or my interlocutors' deontic status, nor there need to be any effects on them, in order for me to speak the truth. Davidson puts this point thus:

> Asserting, commanding, promising, and questioning are not good subjects of investigation if we want to study what is peculiarly linguistic, because in performing such actions we necessarily operate beyond the reach of the conventions of language. Speaking the truth, on the other hand, is specially suited to systematic study just because the conditions for speaking the truth are a matter of linguistic convention. What explains this possibility in part is that speaking the truth is essentially disengaged: it is not done for its own sake, nor, typically, for the sake of any other end; we often do not know it has been done, and do not care. (Perhaps it is obvious on reflection that systematic theory of meaning must be based on a feature of speech behavior that is disengaged in this way.) (Davidson 1970/2020, p. 23)

There is something peculiar about the idea that *speaking the truth* is "essentially disengaged" in the way Davidson is proposing here, especially considering the centrality of truth in our epistemic lives, a point I will come back when considering Davidson's later position in "A Nice Derangement of Epitaphs". For now, let us grant him that and see what follows from it.

If speaking the truth is not determined by its effects, then what determines whether I have done so? As we have seen, the answer is "linguistic conventions". But what type of linguistic convention? Davidson is not very explicit about this, but I agree with Rumfitt (2021) that the only reasonable answer is that the conventions in question are the conventions *specified in the axioms of a truth theory for the language of the speaker*. That is, simplifying a lot, suppose the theory has as axioms that "Olívia" refers to Olívia, that "is smart" is true of a person if, and only if, the person is smart, and that a sentence with the form of a predication is true if, and only if, the predicate is true of the referent of the noun. In that case, I speak

the truth in uttering "Olívia is smart" if, and only if, Olívia is smart.[26] And I speak truly because the axioms capture the *conventions* of referring to Olívia by using "Olívia", of using the predicate "is smart" to characterize smart people, and of using sentences with the form of predications to characterize their subjects.

As a nice consequence of this account, we have a straightforward story about how the theory relates to the speaker's competence. The theory describes the speaker's competence because the speaker *explicitly knows* the theory. Of course, the speaker does not need to know the axiomatic form of the theory. Rather, the speaker knows the relevant conventions, i.e. that "Olívia" is used to refer to Olívia, etc. This is nice because, as Davidson makes clear, it avoids any appeal to tacit or implicit knowledge, or even to the idea that the theory's structure mirrors the structure of the competence of the speaker in any sort of mysterious way:[27]

> This way of showing there is a difference in the semantic structure of [two example sentences] requires no appeal to "the speaker's tacit knowledge" of the grammar or the "intrinsic competence of the idealized native speaker". It rests on the explicit knowledge any speaker of English has of the way in which [the example sentences] may vary in truth under substitutions for the word "John". (Davidson 1970/2020, p. 35)

Interestingly, as remarked by Rumfitt (2021), this provides an answer to a notorious problem with Davidson's account, the so-called *Foster's Problem*.[28] In brief, Foster (1976) objected to Davidson that merely having an extensionally correct truth theory for a language was not enough to characterize the meaning of its sentences, since, given a correct theory that generated, say, as a theorem that "Olívia is smart" is true if, and only if, Olívia is smart, one could construct a new theory that generated as a theorem that "Olívia is smart" is true if, and only if, Olívia is smart and $2 + 2 = 4$. This could be done by tweaking the axioms: whereas in the original axioms we had that "is smart" is true of a person if, and only if, the person is smart, now we would have as an axiom that "is smart" is true of a person if, and only if, the person is smart and $2 + 2 = 4$. Since the old theory was by hypothesis extensionally correct, so is the new theory. Davidson thus needs a criterion to rule out the new theory, and the account sketched above provides a neat answer: the new theory is ruled out because its new axiom does not correspond to any purely linguistic convention. Rather,

[26] I leave as an exercise to the reader to prove this from the axioms; extra points for a more rigorous formalization, perhaps along the lines of Lepore and Ludwig (2005, chap. 4).

[27] Classic developments of these ideas include Evans (1981/1985) and Davies (1981, 1987). Cf. also Fricker (1982-1983) and Antony (1997) for discussion. As mentioned in note 4, Lepore and Ludwig (2005, chap. 9) explicitly defend an attribution of this type of theory to Davidson throughout his career.

[28] Cf. Lepore and Ludwig (2005, chap. 8) for an overview and discussion.

it corresponds to a linguistic convention and an arithmetical truth.[29] Moreover, since, in the John Locke Lectures account, speakers have *explicit knowledge* of those conventions, one can test which theory is to be preferred simply by consulting one's explicit knowledge:

> T-sentences are as easy to judge true or false as anything that could count as evidence for the correctness of a theory of meaning. For, as remarked, they contain no semantic vocabulary (except for the truth predicate), no reference to logical form, no general laws. To recognize them as true, we require no insight into synonymy, deep structure, or meanings. Yet there is nothing more to the question whether a theory of truth is correct than the question whether the T-sentences it entails are true. So it seems that a theory of truth satisfying Convention T, though it yields, if I am right, important and by no means obvious results about a language, can be tested by appeal to knowledge plausibly explicit to a native speaker and as available as any to the radical translator or descriptive linguist. (Davidson 1970/2020, p. 34)

Summing up, then, the theory developed by Davidson, according to which theories of truth are theories about conventions for speaking the truth, pay up some nice dividends. First, since it is based on the idea that speakers have explicit knowledge of these conventions, it allow us a straightforward story about how our semantic theories relate to linguistic competence. Second, because the theory is about linguistic conventions, it also gives us an explanation of the systematic knowledge of the speakers that accounts for its diachronic stability. Third, it gives us a nice criterion to select between extensionally adequate theories, namely, we choose the theory that captures the linguistic conventions of the speaker. Given such riches, why did Davidson abandon this account? The reason is that he eventually came to see that conventions could not play the role he first envisioned for them in his theory. In the next section, I will explore how he came to that conclusion.

3.4 Against Linguistic Conventions: Davidson and "A Nice Derangement of Epitaphs"

In the last section, we saw how Davidson singled out speaking the truth as a peculiar kind of act, whose conditions of success were completely exhausted by purely linguistic conventions. As mentioned in the Introduction, his espousal of this theory was short-lived: already in 1974 he had already given up on the idea of explaining meaning through appeal to linguistic conventions. Although this skepticism was already expressed in his reply

[29] As said, I am following here Rumfitt's line.

to Foster and in the 1984 article, "Communication and Convention", it is certainly in "A Nice Derangement of Epitaphs" that his reasons become more explicit and systematic. In this section, I will lay out his argument for the new position and assess its consequences for Davidson's overall program.

As is well-known, Davidson's argument relies on the phenomena of malapropisms and other deviant speech patterns. The point is clear enough. Suppose Mrs. Malaprop utters a phrase with an unconventional meaning, such as "a nice derangement of epitaphs". We understand her to mean by that a nice arrangement of epithets. The question is, how do we do so? Apparently, in order to understand her, we cannot be relying on conventions, since, if we were, we would understand her to mean a nice derangement of epitaphs, whatever that is. Davidson takes this to show that conventions are neither sufficient nor necessary for linguistic communication, and therefore explanatory idle. Let us examine this claim in detail.

Recall that, whatever else they are, conventions are primarily patterns of behavior. Hence, in order to employ a linguistic convention in interpreting another person's utterance, I am projecting this regularity onto the person's utterance, effectively subsuming the utterance under this pattern of behavior. In other words, it is not enough that a convention be in force for me to understand an utterance in accordance with this convention; I must also *interpret* the utterance *as* falling under the convention. But this opens the space for me to interpret an utterance as *not* falling under the relevant convention *even when* the convention is in force. This is what allows me to interpret "a nice derangement of epitaphs" as meaning a nice arrangement of epithets. What this shows is that, even supposing there are linguistic conventions in force, they are not sufficient for linguistic communication, as *an interpretive effort* on the part of the hearer is also required. This interpretive effort may require much more than knowledge of linguistic convention, as it may have to draw on a theory about who the speaker is—indeed, that is the main point of radical interpretation. In other words, Davidson thinks that, once we have a sufficiently robust theory of interpretation, conventions become explanatory idle, as the theory will cover both the so-called conventional cases and the unconventional ones.

This point is important because, as Stainton (2016, pp. 11ff) makes clear, there is an understanding according to which the non-sufficiency of conventions for linguistic communication is entirely uncontroversial. Stainton mentions the case of context-sensitive expressions, such as demonstratives. In order to understand an expression containing a demonstrative, one needs to appeal to extra-linguistic features, such as the speaker's intention or features of the context to fix the reference of the demonstrative. The linguistic conventions may at best say that, if the utterance of a demonstrative refers to a certain object, then

some truth-condition follows.[30] But they cannot determine the reference of the demonstrative itself, and, since to understand the utterance I need to understand the reference of the demonstrative, mastery of linguistic conventions is not sufficient for understanding utterances containing demonstratives.[31] Stainton observes, however, that Davidson wants to go further than that. The point is not merely that some pragmatic processes are necessary to solve local problems of context-sensitivity. Rather, Davidson is proposing that *even in the cases where there is apparently no context-sensitivity*, our interpretation proceeds using other cues than linguistic conventions—hence his focus on malapropisms.[32]

Before unpacking this further, let us see what we can say about the *necessity* of linguistic conventions for communication. Here, Davidson's example of malapropisms is less helpful, since one could say that, even if linguistic conventions are not sufficient to understand Mrs. Malaprop, they are nevertheless necessary. The argument would go as follows. What makes us understand that, by "a nice derangement of epitaphs", Mrs. Malaprop means a nice arrangement of epithets, is our knowledge that "derangement" is phonetically similar to "arrangement", that "epitaphs" is phonetically similar to "epithets", *and* our knowledge of the linguistic conventions that "epithets" designates epithets and "arrangement" designates arrangements.[33] Let us grant, for the sake of the argument, that this analysis of how we manage to interpret malapropisms is correct. Can we find a better example that shows that linguistic conventions are not necessary?

I believe we can. Consider the case of *homesigns*, the language developed by deaf children of parents that, for one reason or another, do not communicate with their children through a sign language.[34] In those cases, the children are *inventing* a language "from the ground up" (to use the apt expression from the title of Begby's (2017) article), so there is no regularity involved in their gestures. It follows immediately that the interpreter cannot be relying on any regularity to understand these children's linguistic gestures (since there need be no such regularity, at least not at first), and, with no regularity, there is no linguistic convention involved. So linguistic conventions are not necessary for an interpreter to understand the

[30] This model was first proposed by Burge (1974) and later developed by Higginbotham (1988, 2002). Cf. Heck (2014) for a defense of the idea.

[31] Or so it seems. Stojnić (2021) argues that mastery of linguistic conventions *is* sufficient to understand utterances containing demonstratives, because these linguistic conventions also fix the reference of the demonstrative. I unfortunately do not have the space to engage with Stojnić's very interesting work.

[32] Stainton (2016, pp. 14ff) helpfully puts the point in terms of a progressive abandonment of a *code model* of linguistic communication, according to which linguistic expressions are *codes*. I think this is exactly right.

[33] This is the line taken by Lepore and Stone (2017).

[34] For an introduction to this incredible phenomenon, cf. Goldin-Meadow (2003). Although linguists such as Jackendoff (2003, p. 99) have long since used homesigns as evidence for language-specific cognitive devices—such as Universal Grammar—, the only philosophical work that I know of which extracts from this phenomenon broader consequences for the philosophy of language is Begby (2017).

linguistic actions of another person.

How, then, do we manage to interpret another person? Davidson's answer is complex and developed in a series of papers (Davidson 1973/2001b, 1974/2001, 1975/2001, 1980/2004), but the gist of it can be summarized as follows. One of the most important activities in which we engage, for Davidson, is *interpreting* another person, that is, we try to make sense of a person's patterns of actions and behaviors.[35] This activity of interpretation is not unconstrained, though. Crucially, we interpret others as obeying a *constitutional ideal of rationality* (Davidson 1980/2001, pp. 222-223), that is, we interpret their actions as conforming to rational patterns.[36] There are at least two constraints that inform this ideal of rationality. First, we interpret others as being logical and as being largely correct—the principle of charity. Second, we interpret their beliefs and desires as roughly satisfying the principles of Bayesian decision theory.[37] Obviously, that does not mean that the lay person has knowledge of Bayesian decision theory when interpreting another person; it does mean that Bayesian decision theory codifies the kind of principle used by the lay person to interpret another in a rational way.

One important change Davidson introduces to Jeffrey's version of decision theory is to replace the notion of a subject preferring one *proposition* rather than another with the notion of a subject preferring that one *sentence* be true rather than another (Davidson 1980/2004, p. 161). The result of this is that truth conditions for sentences become inextricably intertwined with belief and desires, since these truth conditions *just are*, in a sense, the objects of the subject's preferences. In other words, meaning (in the sense of truth conditions for someone's utterances), belief, and desire cannot be understood in isolation. Instead, "meaning, belief, and desire will be treated as fully coordinate elements in an understanding of action" (Davidson 1980/2004, p. 152).

Each time, then, I interpret another person's utterance, I am considering this speech act of hers as taking place against a larger pattern, defined by this constitutive ideal of rationality. Of course, that does not mean that each utterance is interpreted *ab novo*; on the contrary, I generally come equipped with what Davidson (1980/2004, pp. 100-101) calls a *prior theory* about the meaning of the subject's utterances, in the form of an axiomatized truth theory (I will get to later about what this means exactly). On the other hand, this interconnection between meaning, desire, and belief has as a consequence that one has a certain latitude

[35]That is why interpreters such as Glüer can say that "The interpreter thus is the hero, the main character of Davidson's philosophy" (Glüer 2011, p. 4): it is the activity of interpretation which unifies all the different strands of Davidson's philosophy.

[36]Cf. Rescorla (2013) for an overview.

[37]Davidson (1980/2004, pp. 160ff) favors, with some adjustments, the theory as developed by Jeffrey. The adjustments mainly concern the fact that Jeffrey uses propositions, whereas Davidson, unsurprisingly, prefers to talk about sentences. Cf. Jeffrey (1990) for a very nice textbook treatment of his theory.

when interpreting a person's utterances, especially if the utterance appears deviant in some way. That is, given a deviant utterance, I can either attribute the deviance to the subject's beliefs or the subject's theory of meaning. In the case of Mrs. Malaprop, for instance, I can either attribute to her the belief that she is an expert at nice derangements of epitaphs, or else that she is employing the words "derangement" and "epitaphs" in some non-standard way. If I opt for the later, then I am developing what Davidson called a *passing theory* of her utterance (Davidson 1980/2004, p. 101).

These notions of prior and passing theories are what allows Davidson to bypass the notion of convention entirely. Instead of coming equipped with knowledge of *conventions*, interpreters come equipped with *prior beliefs* (a prior theory) about what each word means. These beliefs can be more or less entrenched, and they form the basis according to which I will interpret another person's utterances. After I hear the person's utterances, I assess the rationality of her utterance against the background of this prior theory and of her behavior more generally. If I find that I make better sense of her utterance by adjusting the meaning of her words, I then revise my prior theory into a passing theory which incorporates these adjustments. Depending on the case, I may either consider your use authoritative, and thus incorporate it into a new prior theory that I will use as a basis for interpreting future utterances, or else I may consider your use as deviant, and only use it to interpret it you, or perhaps you in this particular occasion.[38]

Now, it may very well be that the prior beliefs of many speakers coincide, thus giving rise to a regularity in linguistic behavior that we may wish to call a linguistic convention. Since, however, the order of explanation flows from the prior beliefs to the linguistic conventions, this account is much more tolerant of individual idiosyncrasies than the account which assumes as primary linguistic conventions. That is, whereas the convention-first account implies that there is *uniformity* of meaning, enshrined in the convention, the prior theory account allows for speakers to have slightly (or even, in the case of Mrs. Malaprop, massively) different prior theories. Moreover, Davidson's account allows for the interlocutors in a conversation to *mutually adjust* their theories in response to each other, thus allowing communication to emerge as a real *joint activity* (Clark 1996) in which the participants all contribute to the constitution of meaning, and not as a succession of monologues.[39]

Let us take stock. In this section, I followed Davidson's argument against the explanatory relevance of conventions. In its place, we saw how Davidson developed a more flexible notion of prior and passing theories to account for how people interpret each other against the background of the constitutive ideal of rationality. In the last section, we saw how

[38] Cf. Begby (2017, p. 44), and the references therein, for more on how social statuses may affect this choice as well.

[39] Begby (2017, p. 44n34) also notes this convergence with the literature on discourse collaboration.

Davidson developed an account of semantic theories as codifying speakers' knowledge of linguistic conventions. As mentioned there, this account had three positive consequences. First, it gave us a straightforward story about how our theory related to the linguistic competence of the speakers. Second, it gave us an explanation of the character of the systematic knowledge of the language enjoyed by the speakers. Third, it also gave us a criteria to choose between different but extensionally equivalent theories. Since all those relied on the notion of convention, something needs to be said about if the new theory still retains those theoretical advantages, or if, by prescinding from convention, it also prescinds from these advantages as well. In the next section, I will argue that the new theory is only an improvement over the old theory if, together with the appeal to conventions, we also abandon any hope of using a theory of truth as a systematic semantic theory.

3.5 Assessing the New Theory

In this section, I will examine the three issues in turn:

1. The story the new theory has to tell between truth theories and semantic competence.
2. How the new theory accounts for the systematic character of linguistic competence.
3. Whether the new theory furnishes a criterion for us to choose between extensionally equivalent theories.

3.5.1 Semantic Competence

As I mentioned in the last section, Davidson saw interpretation of a person's utterance as part of a more general interpretation of a person's *actions*. Unsurprisingly, then, I believe his theory of action may shed some light on the relation between the theory of truth and semantic competence. A quick summary of that will be thus helpful.[40]

Ever since his first publication in action theory, the seminal "Actions, Reasons, and Causes" (Davidson 1963/2001), Davidson insisted that reasons are a kind of cause. The idea, at first blush, seems relatively simple. We saw in the last section how Davidson considered the interpretation of a person to be the placing of her actions in a wider rational pattern. In particular, the ascription of beliefs and desires to a person amounts precisely to a kind of *rationalization*: we explain the person's action of ϕ-ing by, for example, citing the person's desire for ψ and her belief that ϕ-ing was the best way to achieve ψ. In that case, we say that the belief and the desire, together, constitute the agent's *reason* for ϕ-ing (supposing the

[40]The account presented here is heavily indebted to Stoecker's (2013) excellent overview.

agent ϕ-ed *because* of the belief and desire, of course). What Davidson wants to say is that this type of explanation is also a *causal* explanation.

At first, Davidson's insistence that rationalizations are causal explanations seems to fit poorly with two of his thesis regarding causation, namely that "where there is causality, there must be a law" and that there cannot be "strict deterministic laws on the basis of which mental events can be predicted and explained (the Anomalism of the Mental)" (Davidson 1980/2001, p. 208).[41] After all, if rationalizations are causal explanations, they presumably are linking belief-desire pairs with action by way of causation. Thus, as per the first principle above, there must be a causal law covering this particular belief-desire pair with that particular action. But, by the Anomalism of the Mental thesis, there can be no strict deterministic laws covering mental events such as the onset of a particular belief-desire pair. So there can be no causal explanation linking belief-desire pairs with actions, contrary to our assumption. What has gone wrong?

There are a couple of threads here to untangle.[42] First, there is a distinction to be made between two uses of "cause", one to describe a relation between events, the other to *explain* one phenomenon in terms of another. Second, there is a distinction between strict causal laws and the causal generalizations which underlie causal explanations. Finally, there is the notion of a causal power and its role in causal explanations. Let us tackle each of these issues in turn.

Regarding the first issue, Davidson accepts that "causes are individual events, and causal relations hold between events" (Davidson 1967/2001, p. 161). That is the primary use of "cause", then, as a two-place predicate relating names for individual events. Beliefs and desires are not events, however, but states, so if reasons are typically thought of as belief-desire pairs, then "reasons are not causes" (Davidson 1993, p. 288). But causal *explanations* go beyond causal *statements*, since explanations are typically *interest relative* in a way that simple causal statements are not:

> Everybody allows that most talk about causality is interest-relative; what we call "the" cause of some event is some feature chosen from the totality of causal factors which particularly interests us, something we find surprising or out of the ordinary. In giving a causal explanation of an event we normally take for granted a great deal of background; what we typically want to know is what to add to that background to make the occurrence of the effect intelligible. (Davidson 1993, p. 287)

[41] Cf. Stoecker (2013, pp. 23ff) for commentary.

[42] Again, I am strictly following Stoecker's exegesis here. Stoecker's article is very useful because Davidson's thought developed over the course of numerous essays, so that it may be difficult to see how systematic it is just by reading the individual articles.

So, on the one hand, we have causal statements which relate events as cause and effect, and, on the other hand, we have causal explanations that make salient one causal factor among many for the purpose of making intelligible a given effect. Causal statements are *extensional*, in that they are true regardless of the description we use to pick out the relevant events. For instance, if my turning on the lights caused the burglar to flee, this relation between my turning on the light and the fleeing of the burglar remains true if I re-describe the relevant events as the movement of my fingers and the movement of the burglar, i.e. if say that the movement of my fingers caused the movement of the burglar's legs. Importantly, though, how I choose to describe the events may affect the explanatory power of the statement, e.g., while it is true that the cause of the fleeing of the burglar caused the fleeing of the burglar, this is utterly uninformative, and so not a causal *explanation*. Causal *explanations*, then, are *intensional*, since one description of the events may pick out precisely the relevant aspect of the event which was crucial to render it intelligible, where which aspect is relevant is relative to the background of the inquiry at hand.

Notice how this account of causal explanations dovetails nicely with Davidson's account of *interpretation*. Recall from last section that to interpret another person is to place that person's actions in a larger pattern of rational behavior. In particular, to interpret a person's particular actions is to attribute a pair of belief-desire such that the action makes rational sense, i.e. it is made intelligible against the background of the person's behavior. As should be clear, to interpret another is *not* to describe some kind of strictly deterministic causal law according to which her behavior follows inexorably from the fact that she has particular beliefs and desires. Indeed, the attribution of the belief-desire pair to the person only explains the action *insofar as* the background of the person's behavior, and the theory of rationality involved, is in place—another person with the same belief-desire pair could have acted completely different in the same circumstances. So why doesn't this contradict Davidson's insistence that "where there is causality present, there must be a law"?

Here, the distinction between causal statements and causal explanations becomes relevant. Causal explanations need not be causal statements. Consider the following scenario. I am intrigued as to how a match from a set of wet matches could light up when stricken. Someone then explains to me: "This particular match was dry". This is a causal explanation, since it makes intelligible to me how the match could light up when stricken. But it is not a causal statement, since it is not relating events—the dryness of the match is not an event, but a *state* of the match. Similarly, when I mention a particular belief-desire pair as a causal explanation for a certain action, I am not making a causal statement, but merely pointing out how the relevant state of the person contributed to her behavior. That is not to say that there are no events involved—in fact, Davidson insists that there will *always* be events involved: "The 'total' cause of such an event must always include an event or change,

whatever we happen to mention as 'the' cause" (Davidson 1993, p. 287). It may be that, like in the match case, the events are particularly salient—in the match case, the cause event is the striking of the match, in the action case, it may be a perceiving, a coming to believe, or something similar. Or it may be that we are ignorant of the event in question. In any case, the causal law will be a law that may apply to the events involved in the causal explanation, but not necessarily all the causal factors mentioned will be events. The point is that Davidson requires that every causal explanation involves a causal relation between events, and every causal relation between events requires a law, but we may give a causal explanation without knowing the events involved, let alone the law relating them:

> The reconciliation [between the Humean and the anti-Humean about causality] depends, of course, on the distinction between knowing there is a law 'covering' two events and knowing what the law is: in my view, Ducasse is right that singular causal statements entail no law; Hume is right that they entail there is a law. (Davidson 1967/2001, p. 160)

At this point, another question naturally arises: if causal explanations do not cite any strictly deterministic laws, how can they still be explanatory? Davidson makes use here of the distinction between two types of explanation:

> In my opinion Honderich has failed to note the difference between events described in terms that allow the application of laws without ceteris paribus clauses, laws that make no use of causal tendencies, potentialities, or dispositions, and laws that, by using such devices, allow us to choose what we call the cause according to our special explanatory interests. (...) Laws of these different sorts all yield explanations, but explanations of different sorts. (Davidson 1987/2004, p. 113)

Let us reserve the term "law" for the first sort of explanation from the above quotation—those that involve strictly deterministic causal laws—and, following Stoecker (2013, p. 23), call the second sort of explanation a *causal generalization*. This is an important distinction, and points towards the Quinean roots of Davidson's philosophy.[43] Recall that, for Quine, the use of the dispositions idiom is considered as a sort of *promissory note*: we use the dispositions idiom to keep track of patterns of causal relations *that we do not know yet how to explain*, but that we suspect will eventually yield to science. Davidson extends this to explanations that explicitly mention causality:

[43]I am thinking specifically here of Quine's treatment of the disposition idiom. Cf. Quine (1969b, 1973, Chap. 1).

> Unavoidable mention of causality is a cloak for ignorance; we must appeal to the notion of cause when we lack detailed and accurate laws. In the analysis of action, mention of causality takes up some of the slack between analysis and science. (Davidson 1973/2001a, p. 80)

Importantly, some concepts, like the concept of *sunburn*, have a built-in reference to causality: to be sunburned is to have one's skin burned by the sun. Following Davidson, let us call these type of concepts *causal concepts* (Davidson 1990/2004, p. 95); they are the concepts we usually appeal to in our causal generalizations. They are like Quine's dispositions idiom, placeholders that we use until we have more precise knowledge of the mechanisms behind the phenomena they track. To use them is not to forego explanation, though, since in employing them in a causal explanation we are explicitly saying that there is some underlying state, tracked by the concept, and which has caused this particular event. As Davidson says, even the claim that this pill has put someone to sleep because of its dormitive power (à la Molière) still "has content, since the pill might have put the person to sleep not because it was soporific but because under the special circumstances that obtained it acted as a placebo" (Davidson 1990/2004, p. 95). For Davidson, mental concepts are all causal concepts:

> The propositional attitudes, the semantics of spoken words, and behavior as we normally understand it, are all like this [i.e. are all causal concepts]. The reason, both in the case of the attitudes and in the case of semantics, is the same: what our words mean, and what our thoughts are about, is partly determined by the history of their acquisition. (Davidson 1995/2004, p. 121)

The resulting picture is this. Mental concepts, including semantic concepts, are used in causal generalizations to keep track of underlying physical regularities or mechanisms. Rationalizations, then, are explanations insofar as they are instances of these causal generalizations, and their explanatory power results from the postulation of an underlying mechanism, thus excluding other possible causes for the person's behavior. Now, if positing underlying mechanisms was all there was to rationalizations, then they would be explanatory on a par with positing dormitive powers to sleeping pills. Sure, like Davidson says, such talk is not *wholly* empty, but it is also not terribly informative. What gives the explanatory purchase to our rationalizations is that they impose a *structure* on the subject's dispositions.

To explain this point, Davidson introduces an analogy with measurement. When measuring weights, for instance, we attribute to objects certain numbers, but what is important is not the exact numbers we assign to the objects (hence why we can measure weights in pounds or kilos), but the *relationships* between these numbers, so that, e.g., an object with twice the weight of another will be assigned twice the number assigned to this latter object.

Similarly, when assigning beliefs and desires to a subject, the *relationships* (e.g. entailment relations among beliefs) among these attitudes matters as much as their content:

> Just as in measuring weight we need a collection of entities which have a structure in which we can reflect the relations between weighty objects, so in attributing states of belief (and other prepositional attitudes) we need a collection of entities related in ways that will allow us to keep track of the relevant properties of and relations among the various psychological states. (Davidson 1989/2001, p. 60)

We can now apply these ideas to the semantic theories proposed by Davidson. When we describe a person's prior or passing theory by using an axiomatic truth theory, we are using the axioms of the truth theory to keep track of the underlying dispositions of the person. As per the above, these dispositions are ultimately to be cashed out in terms of physical mechanisms instantiating strictly deterministic laws, though it is doubtful whether we will ever achieve this level of knowledge. Moreover, the axioms work not just by picking out dispositions one by one, but by attributing significant structure on the subject's dispositions, so that, for instance, to lose the disposition specified by a lexical axiom is to lose the capacity to understand *any* sentence in which that lexical item appears.[44] Notice that the person whose dispositions are being tracked by the truth theory does not need to *know* the theory in order for the theory to do its job. Rather, the theory is a *tool* for us, theorists, to systematize—rationalize, interpret—the person's behavior. As long as the theory is keeping track of real physical mechanisms in the person, it will be doing its job, regardless of whether or not the person has explicit knowledge of the theory (indeed, the person most likely will *not* have explicit knowledge of the theory, unless she is a semanticist).

This creates a problem, however. We saw in the last section that Davidson accounted for people's ability to communicate in terms of prior and passing theories. What is the status of *these* theories? Clearly, they cannot be merely a tool for the *semanticist*, since non-semanticists employ them all the time in order to communicate. Notice that the problem is not that people merely employ prior theories in their communication—this could be readily accounted for in terms of their dispositions, as per above. Rather, the problem is that they *project* those theories to their interlocutors, and thus have *beliefs* about the prior and passing theories of each other. If they have beliefs about these theories, they cannot be merely tools for systematizing the person's state. On the contrary, not only must they have some kind of psychological reality, but they must be *accessible* to the speakers. In other words, we are back to the idea that knowledge of (or belief in) the theory must be *explicit*.

[44] As is well known, this point was heavily emphasized by Gareth Evans (1981/1985).

It is instructive to compare Davidson's position, as described above, with Chomsky's, since this problem does not arise for Chomsky.[45] By the time of his review of Skinner's *Verbal Behavior* (Chomsky 1959, p. 57) and of the first chapter of *Aspects of a Theory of Syntax* (Chomsky 1965), Chomsky was comparing the internal grammar to a kind of deductive theory internalized by the speaker:

> It is not easy to accept the view that a child is capable of constructing an extremely complex mechanism for generating a set of sentences, some of which he has heard, or that an adult can instantaneously determine whether (and if so, how) a particular item is generated by this mechanism, which has many of the properties of an abstract deductive theory. Yet this appears to be a fair description of the performance of the speaker, listener, and learner. (Chomsky 1959, p. 57)

In his first published book, *Syntactic Structures*, Chomsky (1957) had already proposed that a(n abstract) grammar should be understood as a kind of formal system, with a primitive vocabulary and formal rules for deriving complex expressions from the set of primitives, similar to how an abstract deductive theory derives theorems from its axioms and rules of inference.[46] By then, the psychological hypothesis that speakers somehow had this formal system internalized was not yet explicit;[47] nevertheless, it is clear that when, in the quotation above, Chomsky describes the internal mechanism by which a speaker generates a language as having "many of the properties of an abstract deductive theory", it is to this model that he is alluding. Of course, this does not require that speakers have some kind of *explicit* knowledge of the inner workings of these mechanisms:

> This is not to say that he [the speaker] is aware of the rules of the grammar or even that he can become aware of them, or that his statements about his intuitive knowledge of the language are necessarily accurate. Any interesting generative grammar will be dealing, for the most part, with mental processes that are far

[45] Davidson (1970/2001, 1970/2020, p. 34) himself employs this contrasts in his early phase, when he still thought speakers had explicit knowledge of an axiomatic theory in the form of knowledge of conventions. What is interesting is that, according to the story told here, Davidson's position evolved to something similar to Chomsky's, except that his object, a theory of truth, resisted this position, as we will below.

[46] Chomsky was basically modeling a grammar as a Post Production System. Cf. Chomsky (1956) for more details on the idea, Smullyan (1961, chap. 1) for a very elegant presentation of Post Production Systems, and Pullum (2011) for a very thorough (and critical) exposition of the mathematical content of Chomsky's early works.

[47] Which, as Collins (2008, pp. 22-24) argues, may have caused some misunderstandings. Cf. Newmeyer (1986, chap. 2) for how many of those who were at first enthusiasts of *Syntactic Structures* (such as Charles Hockett) were later appalled by the explicitly cognitivist turn of *Aspects*.

> beyond the level of actual or even potential consciousness; furthermore, it is quite apparent that a speaker's reports and viewpoints about his behavior and his competence may be in error. Thus a generative grammar attempts to specify what the speaker actually knows, not what he may report about his knowledge. (Chomsky 1965, p. 8)

The last sentence of the above quotation may give the impression that, irrespective of its conscious status, the grammar internalized by the speaker consists in rules for the manipulation of internal representations—such rules constituting this so-called "knowledge" of the grammar.[48] This sentence could then be used to assimilate this early Chomskyan idea to the so-called "Theory Theory" of cognitive psychologist—compare the above quotation with the following one by Alison Gopnik, one of the foremost defenders of the Theory Theory:

> The basic idea is that children develop their everyday knowledge of the world by using the same cognitive devices that adults use in science. In particular, children develop abstract, coherent systems of entities and rules, particularly causal entities and rules. That is, they develop theories. (Gopnik 2003, p. 240)

Chomsky, however, *rejected* the Theory Theory, at least in the domain of syntactic knowledge (Chomsky 2003a). His reason for this rejection is instructive: there is, for him, a clear disanalogy between the syntactic systems developed by children and scientific theories.[49] Granted, says Chomsky, that children and scientists both form, respectively, their syntactic systems and their scientific theories by employing unconscious theory-formation mechanisms. It does not follow that the output of these mechanisms is the same in both cases. In particular, the syntactic systems developed by children are *not* accessible to consciousness in the way that scientific theories *are* accessible to the scientists who developed them. Indeed, given that these syntactic systems are just the fully developed state of the language faculty, there is a strong sense in which they are not any kind of *knowledge* at all, implicit or otherwise, for knowledge implies the possibility of mistake, and there is no such a possibility in the case at hand.[50]

[48] Whether this impression is correct for the Chomsky of *Aspects* is a controversial matter. Cf. Collins and Rey (2021) for details. Although interesting, this historical matter is not relevant for my comparison with Davidson.

[49] Since my aim here is just to clarify Chomsky's position, I will not be able to do full justice to Gopnik's views. But do see Fuller (2013) for detailed criticism of Gopnik that is much in the same vein as the view developed here. Incidentally, an aspect of Chomsky's view that I will not touch here is his notion of an innate science forming faculty; cf. Collins (2002) for a critical examination of this notion.

[50] Alexiadou and Lohndal (2021, p. 46n4) make much the same point: "However, when it comes to each person's internal grammar, the question of truth does not make sense."

This last point bears emphasizing. Whereas early Chomsky, as we have seen, emphasized the fact that the syntactic system developed by the child resembled a theory, starting with the *Lectures on Government and Binding* (Chomsky 1981/1993), he began to emphasize more and more the biological aspect of this system. In particular, according to the Principles and Parameters framework, the child is born with a kind of *organ*, the language faculty, whose development is constrained by certain principles and which realizes certain parameters depending on the context in which it develops. That is, the language faculty has a certain architecture, and the developmental path of this faculty is basically a matter of "selecting" which values of the parameters to realize. As Chomsky would later put it, this rough idea describes "the 'process of learning', though 'growth' might be an appropriate term" (Chomsky 2018, p. 28, footnote suppressed). One can then think of the external input as merely triggering certain "switches" corresponding to the parameters.

The upshot is that the syntactic system described by the linguist should not be thought of as a system of *propositions*, say, as a kind of rules database that is consulted by a parser. Indeed, commenting on the similar idea that in acquiring a language a person acquires a system of propositional attitudes, Chomsky says that he has "never meant anything of the sort", agreeing "with Fodor that the idea makes little sense" (Chomsky 2018, p. 34). Rather, the person *develops* an organ whose states obey the general principles outlined by the theory without representing them.[51] Obviously, the theory itself is propositional, but that does not mean that the object described by the theory is propositional, anymore than the fact that the theory of gravitation is a system of propositions implies that gravity is somehow propositional. Curiously, Gopnik presents this as an *objection* to Chomsky:

> But even if theory formation can't explain syntax for this reason, this outcome should not be an altogether happy one for Chomskian theory. For the same arguments that say that syntactic knowledge is not a kind of theory call into question whether it is really a kind of knowledge either. (Gopnik 2003, p. 251)

It should be clear by the above that Chomsky embraces the consequences presented in Gopnik's last sentence: syntactic "knowledge" is not a theory and it is not a kind of knowledge, if by the latter we mean a propositional attitude.[52] Notice that this has as a consequence that, though the syntactic theory is of the speaker's internal states, such states need not represent anything in order for them to be linguistic states.[53] We use the theory

[51] Cf. Allott and Smith (2021, p. 534).

[52] To be sure, Chomsky does protest against construing knowledge solely in terms of propositional attitudes—cf. Chomsky (2018, p. 35). But the point is that *in the technical sense* meant, we do not have knowledge of the syntax of our I-language, since there is no such knowledge to be had.

[53] This point is argued forcefully by Collins (2004, 2014).

to *keep track* of the structural relations among the states of the speaker and of the general architecture of the language faculty, but the states are not necessarily representing anything.

It is this last feature which brings Chomsky close to Davidson. We saw above that Davidson thought of our semantic theories (and our attribution of propositional attitudes) as ways of keeping track of the structural relations among the states of a person, and more specifically to keep track of the causal basis of such states. There is a crucial difference, though. Chomsky does not think the speaker or the hearer *explicitly* employ the linguistic theory in order to communicate, whereas Davidson's model of communication *does* imply that speakers use truth theories, in the form of prior and passing theories, in order to communicate. As we saw, Chomsky holds that what the theory is actually described is the person's states, and these need not be representational. Davidson clearly wants to make the same move, which is why he condemns "the identification of the object used to characterize a state of mind with an object that the mind 'knows' or is 'acquainted with', and 'object of thought'" (Davidson 1989/2001, p. 64). If, however, the person is to employ the theory to understand another, then this theory must be somehow present to the person's mind—it must be an object of thought.

The problem is that Davidson wants a theory of truth for a language to do two jobs at once. First, it must be a theoretical account of the speaker's linguistic competence—and this is the job it shares with Chomsky's syntactic theories. But second, he also wants the theory to be *consciously used* by speakers in interpreting one another—and this is the job Chomsky explicitly *rejects*, since the syntactic system is *not* consciously used by speakers in interpreting one another.[54] And the theory needs to be *consciously* used, since speakers are *reasoning* about each other's prior and passing theories when communicating. The result is Davidson's ambivalence regarding the status of his own semantic theories, and all the difficulties attending the idea of "tacit knowledge", which are a consequence of this ambivalence.

Unsurprisingly, my proposal is to have two different things doing those different jobs. On the one hand, we have an account of the speaker's linguistic competence which does not appeal to truth conditions. On the other hand, we have an account of how speakers attribute truth conditions to each other's utterances. Once we have clearly separated these issues, we can see how to apply the Davidson-Chomsky idea that our linguistic theory is used to keep track of states of the speakers without attributing them any kind of propositional attitude regarding these states *and* also that speakers use prior and passing theories to interpret each other. Before going into that, however, let us examine the problem of systematic knowledge, since it also bears on these issues.

[54] Of course, the syntactic theory may be used as a parser by a subsystem of the person's mind, but this is a completely different matter.

3.5.2 Systematic Knowledge

When discussing Davidson's John Locke Lectures, I mentioned how one of the virtues of his approach then was that it allowed for an account of our *systematic* capacity to understand novel sentences. This is not trivial, since it is an important requirement on semantic theories that they explain this systematic capacity. Indeed, one need not agree with Davidson's particular implementation of a semantic theory to see the virtue of this general form of explanation; Katz and Fodor (1963/1964, p. 483ff), who had a very different semantic proposal than Davidson's, also explicitly require that a semantic theory explain the creativity of speakers in terms of systematic stable capacities.

Recall that the conventionalist proposal of the John Locke Lectures also provided Davidson with a rather straightforward story to tell. Speakers' semantic competence consisted in knowledge of conventions. When a speaker comes upon a novel utterance, this utterance is presumably made up of familiar words in familiar modes of combination, so she applies her conventional knowledge to the utterances in question, and thus is able to decode it. What allows the application of her semantic capacity to novel utterances is then that this capacity is *stable* and that the utterances of her speech community follow certain *regular patterns*, so that she can project the meaning of the novel utterances from her knowledge. As Davidson points out, this very strength turns out to be this model's downfall, since it fails to account for novel or deviant uses of old words.

On the other hand, Davidson's new account in terms of prior and passing theories appears to have the opposite defect: while it handles novel or deviant uses of words well, it does not explain how our semantic capacity is *stable* and thus systematic. The problem centers on our supposed ability to come up with passing theories on the spot. About this ability, Davidson says:

> A passing theory really is like a theory at least in this, that it is derived by wit, luck, and wisdom from a private vocabulary and grammar, knowledge of the ways people get their point across, and rules of thumb for figuring out what deviations from the dictionary are most likely. There is no more chance of regularising, or teaching, this process than there is of regularising or teaching the process of creating new theories to cope with new data in any field—for that is what this process involves. (Davidson 1986/2005, p. 107)

In this passage, Davidson comes close to saying that our interpretation of novel utterances is *not*, after all, a systematic capacity, since, as he says, there is no chance of "regularising" this process. He does not quite say it, because he still thinks that his *meaning holism* imposes a recursive structure on the passing theories:

> Why should a passing theory be called a theory at all? For the sort of theory we have in mind is, in its formal structure, suited to be the theory for an entire language, even though its expected field of application is vanishingly small. The answer is that when a word or phrase temporarily or locally takes over the role of some other word or phrase (as treated in a prior theory, perhaps), the entire burden of that role, with all its implications for logical relations to other words, phrases, and sentences, must be carried along by the passing theory. Someone who grasps the fact that Mrs Malaprop means 'epithet' when she says 'epitaph' must give 'epithet' all the powers 'epitaph' has for many other people. Only a full recursive theory can do justice to these powers. These remarks do not depend on supposing Mrs Malaprop will always make this 'mistake'; once is enough to summon up a passing theory assigning a new role to 'epitaph'. (Davidson 1986/2005, p. 103)

This seems weak. If people have the capacity to interpret novel utterances without relying on rules, why assume that meaning is systematic? Recall that systematicity, or the supposed fact that repeatable features have the same semantic contribution in each sentence that they appear, was postulated precisely in order to account for the learnability of languages, that is, for the fact that people can understand an unbounded number of sentences on the basis of a finite capacity. Now, however, Davidson is saying that people can interpret novel utterances without relying on any kind of rule at all. Indeed, it seems that he is now affirming precisely what he denied in "Theories of Meaning and Learnable Languages", namely that we do have a capacity to "intuit the meanings of sentences on no rule at all" (Davidson 1965/2001, p. 9). If that is so, then, perhaps, in one utterance, Mrs. Malaprop means epithets by "epitaphs", whereas in the very next utterance she means epitaphs by "epitaphs", so that repeatable features do *not* have the same semantic contribution in each sentence that they appear.

But then, what is the use of positing a specific semantic capacity at all? It seems, rather, that what we have is a *general* capacity for interpreting people, one which allows us to "intuit" the meaning of novel utterances as our passing theories. Elisabeth Camp (2016, p. 120) has forcefully argued this point, turning Davidson's own words against him. As Camp reminds us, in "What Metaphors Mean", Davidson argues against the idea that metaphorical expressions have a metaphorical meaning beyond their literal meaning by saying that to posit such a meaning is "like explaining why a pill puts you to sleep by saying it has a dormative power". Moreover, metaphorical meanings lack explanatory power precisely because, like passing theories, they cannot "be assigned to words and sentences apart from particular contexts of use" (Davidson 1978/2001, p. 247). So why is explaining a person's ability to interpret another by the use of passing theories not an instance of vacuous, "dormative

power" explanation? Note that this problem is exacerbated if we interpret the axioms of the passing theory as tracking causal dispositions of the speaker. In that case, we are literally attributing a one-shot, context-bound causal disposition whose only role is to account for the production or interpretation of *one* sentence—i.e. a completely *ad hoc* capacity. Something has gone wrong.

Let us take a closer look at the process of interpretation. In the typical case, a person hears a certain expression being uttered, retrieves a body of information[55] associated with the expression, and uses (part of) that body of information in interpreting what the other person is saying. We thus have an association between an expression and a body of information, or, if the expression is polysemous, between an expression and several bodies of information. Moreover, since the person draws upon these bodies of information on multiple occasions, they have to be stored in long-term memory. Let us call these bodies of information stored in long-term memory and which underwrite speech production and interpretation a person's *conceptions*.[56] In the most general case, then, we have an association between an expression and a family of conceptions.

A couple of points about conceptions. First, note that each individual may have his own conception associated with a word. So, when hearing the word "cat", a person may retrieve a conception which includes the piece of information that orange cats are typically male, whereas another person may retrieve a conception which includes no such piece of information. This precludes the association between an expression and a (family of) conception(s) being a convention, since there is no social regularity involved.[57] Second, since the person may add pieces of information to her conception, it is not just the content of the conception which individuates the conception, but also its *psychological continuity* for the individual.[58] That is, we should treat conceptions as "individuals", in the sense that they survive changes

[55]For ease of exposition, I am using here "information" in the non-factive sense, i.e. to cover both information and misinformation.

[56]The reader familiar with Machery (2009) will notice that I have hijacked his definition of concepts. I choose to call such bodies of information "conceptions" in order to avoid needless controversy about whether they fulfill all the functions typically associated with concepts. There are two differences between my definition and Machery's that are worth mentioning: (1) Machery's definition is in one sense broader, since he is considering bodies of information that serve multiple cognitive processes, whereas I am restricting myself to linguistic processes, since that is my focus here; (2) in another sense, his definition is narrower, since he only includes those bodies of information that are retrieved *by default* (Machery 2009, p. 11), thus distinguishing the bodies of knowledge that form parts of conceptions from what he calls background knowledge. But, as Begby (2016) argues, Davidson's argument in "A Nice Derangement of Epitaphs" calls into question this latter distinction, so I prefer to remain agnostic here on this question.

[57]The association is certainly *arbitrary*, though. As Davidson (1984/2001, p. 265) puts it when arguing against Lewis, though, "while the conventional is in some sense arbitrary, what is arbitrary is not necessarily conventional". Chomsky (2003b, p. 311ff) makes much the same point against Millikan.

[58]This point was made by Machery (2010, p. 431-432) in his reply to Christopher Hill.

in their parts.[59]

This allows us to divide the process of interpretation into two parts. First, there is the association between the expression and the conception. Since the conception is individuated by its psychological continuity, this association is *stable*, as it survives the updating of the conception. Since this part is stable, it allows for cross-contextual application, and hence for systematic treatment. Second, there is the use of the conception in interpreting the other person's utterance. Since it is obviously dependent on the context, *this* part is not stable, so that the final interpretation is "derived by wit, luck, and wisdom from a private vocabulary and grammar, knowledge of the ways people get their point across, and rules of thumb for figuring out deviations from the dictionary" (Davidson 1986/2005, p. 107). That is not to say that the interpretation process is completely anarchic. Davidson himself laid the groundwork for giving a systematic account at least of its general outline with his theory of radical interpretation, and surely more can be said about this (I will come back to it in the next subsection).

Notice that conceptions do not contribute directly to *truth conditions*. Rather, the interpreter *uses* the conception to arrive at the truth conditions of utterances. *How* the interpreter uses the conception to arrive at the truth conditions of utterances is not amenable to systematic treatment. If we identify the semantic value of an expression with what can be systematically studied by linguistics, it follows that the semantic value of an expression is *not* its contribution to the truth values of the utterances in which it appears, since there is not anything as *the* contribution of an expression to the truth conditions of the utterances in which it appears. That is because there is no rule for determining whichever piece of information belonging to the conception is used at each occasion, so no rule for determining the specific contribution of the conception to the determination of the truth conditions of the utterances. Indeed, in the case the passing theory is incorporated into the prior theory, the utterance may itself force the interpreter to *revise* the conception, updating it with new pieces of information required to make the utterance true.

If conceptions do not contribute to truth conditions, then neither do expressions. But then, what *is* the systematic contribution of an expression to the sentences in which it appears? I submit that it is the only thing stable enough to do this work, namely the conception itself. It may not sound much, to say that the systematic contribution of "dog" is the person's conception of dogs. But it is probably all we can say about a specific open class expression's semantic value. Things only get more interesting when we investigate the *modes of combination* of *types* of expressions, as well as closed class expressions. In other words, when we

[59] I leave the mereological details to the metaphysicians. Perhaps conceptions should be treated as mereological sums of their time-slices, perhaps they should be thought of as having temporal parts, or perhaps some other treatment that does not involve classical extensional mereology is called for.

study the *syntactic* contribution of expressions.

3.5.3 Foster's Problem

Finally, let us tackle Foster's Problem. Recall that the problem was of how to decide between two extensionally equivalent truth theories for the same language—or, better, for the same speaker, since, as I argued when discussing Davidson's "A Nice Derangement of Epitaphs", public languages are not the real target of the theory. Notice, however, that the problem has now changed character. When we considered a truth theory to be a theory for a language, the problem was how to choose between different accounts of the stable features which generated the same truth conditions for the utterances of the language. But now, there are no such features, so the problem is not a *theoretical* problem anymore.[60] Instead, it may be considered as a *practical* problem for the interpreter.

As Davidson says in "A Nice Derangement of Epitaphs", there are no strict rules for an interpreter to decide between competing theories. That is not say this is not a rational process. On the contrary, as our discussion of Davidson's radical interpretation procedure showed, he believes that interpreting another is to make sense of that person's actions as fitting in a rational pattern, that is, as being intelligible against the backdrop of the constitutive ideal of rationality. This provides some needed constraints on interpretation, since the interpreter must construe the speaker as abiding to the canons of rational choice theory, for instance. Still, this may leave room for plenty of indetermination. Is there anything else to constrain the formulation of an interpretation? I believe there is, namely, the *speaker herself*. Let me elaborate on that.

As Herbert Clark (1996) repeatedly emphasizes in his book, communication is a *joint* action. This means that speaker and interpreter are *collaborating* on how to best construe the speaker's meaning[61]. As he puts it:

> In the view I will argue for, the notion "what the speaker means" is replaced by "what the speaker is to be taken to mean." The change is small, but radical. The idea is that speakers and addressees try to create a joint construal of what the speaker is to be taken to mean. Such a construal represents not what the speaker means per se—which can change in the very process of communicating—but what the participants *mutually take* the speaker as meaning, what they *deem* the speaker to mean (...). (Clark 1996, p. 212)

[60] The problem of choosing among competing syntactic theories for a speaker receives the same answer as the problem of choosing among competing empirical theories in general, that is, one weights all the evidence, checks the theoretical virtues and defects (e.g. elegance, simplicity, etc.), and makes an informed choice. Cf. Collins (2010) for discussion.

[61] Clark (1996, p. 212) calls this the principle of *joint construal*.

The tradition from Quine, Lewis, and Davidson often emphasizes what one may call *radical* situations, that is, situations in which the means for communication between parties are extremely meager, because they may belong to alien cultures. This may seem like the task facing the translator or interpreter is unilateral, as if the translator or the interpreter were watching a tape of a person speaking in an alien tongue, trying to decipher what she means. This is especially true of Lewis's (1974/1983) presentation, since he presents the situation almost like a puzzle the interpreter has to solve. Quine (1960/2013, chap. 2) and Davidson (1973/2001b), on the other hand, repeatedly highlight the fact that one must *query* the person, *present* her with novel situations, etc., in order to home in on the correct interpretation. That is, even in this impoverished communicative environment, the interpreter must *work with the speaker* to arrive at the correct interpretation of her utterance.

Even with last paragraph's amendment, the "radical situation" tradition may still give a distorted picture of the communicative situation. Given the impoverished communicative means at hand and the fact that these scenarios typically emphasize just one party trying to understand the other, it may seem that all the collaboration is merely to *clarify* a meaning that was always already available once the speaker made her utterance. As Clark makes clear in the quotation above, however, that is not the case. The collaborative endeavor is also an endeavor to *jointly construe* what the utterance meant. That is because the speaker herself may have no clear communicative intention *before* engaging with her addressee, in such a way that the addressee actively contributes in determining this intention, and therefore the meaning of the utterance itself.

This allows us to contend with a final criticism of Davidson's proposal in "A Nice Derangement of Epitaphs". Objecting to Davidson's account, Lepore and Stone say:

> To confirm that malapropisms are genuine cases of improvised communication, we need to be sure that the speaker has a particular meaning in mind, that the hearer interprets the utterance with the same meaning, and that what enables this success is the interlocutors' nonce insights into each other's mental states, rather than a shared regularity they have joint access to. If interlocutors do not arrive at the same meaning, they have not communicated; they have at most influenced one another. If they rely on a shared regularity to do so, their communication is not improvised; it remains an outgrowth of their antecedent knowledge of the rules of language. (Lepore and Stone 2017, p. 247)

Lepore and Stone seem to argue like this. Speakers have a certain communicative intention; this communicative intention is a private state of mind, which the interpreter has no access to independently of language; so the interpreter must rely on language to determine the intention of the speaker; in particular, she must rely on linguistic regularities in order

to home in on the speaker's intentions. As we saw, part of the problem with this reasoning is that it relies on the supposed fact that the interpreter must arrive *on her own* at what are the speaker's intentions, when in most cases she will need assistance from the speaker. But second, there is a deeper problem with their model, namely the assumption that intentions are *private states of the mind*. This is a huge topic, which I will only be able to lightly touch upon here. But suffice to say that Anscombe (1963) has provided us with another model for characterizing intentions, one that does not rely primarily on a person's private states of mind.[62] As John Schwenkler very aptly summarizes:

> In contrast to this [viz. the idea that intentions are interior acts], the starting-point for Anscombe's inquiry is the way that the concept of intention is put to work in our description of intentional *action*. The first question that she wants us to ask of an agent is not "What does he intend?" but rather "What is this man *doing*?" (cf. §23, 37:3)—and then the concept of intention is deployed as a way to describe *what is happening* in a way that displays the particular form of unity found in human action, and the means–end order embodied in the series of descriptions we can give of it. In this context we use the concept of intention to make *sense* of an action, to understand what connects a person's behavior in its immediate descriptions to the further ends that her movements are meant to serve. And we cannot do this successfully without representing the action as intentional under those descriptions that give the *means* by which the agent's movements are intelligibly related to her ends. (Schwenkler 2019, p. 89)

In other words, to discern someone's intention is not necessarily to discern a private mental state, but rather to come to understand how her actions relate to her purposes or ends. Thus, to understand why Mrs. Malaprop means a nice arrangement of epithets by "a nice derangement of epitaphs", one must understand that her purpose in using the expression "a nice derangement of epitaphs" is to defend herself against the charge that she "deck[s] her dull chat with hard words which she don't [*sic*] understand" (Sheridan 1798, p. 49), and then reason that for her to say she is good at producing "a nice derangement of epitaphs" does not serve this purpose, so she must have meant something else. Knowing that Mrs. Malaprop frequently confuses words similar in sound, one then arrives at the correct interpretation.

Or, to examine a more mundane example, consider the following exchange, taken from Clark (1996, p. 144). A person asks another, "I think I am parked in an illegal parking zone", to which the other replies "There is a garage round the corner". In British English, apparently (I am taking Clark's word for this here), "garage" is ambiguous between *gas*

[62] I am grateful to Eduardo Marchesan for bringing to my attention the relevance of Anscombe in this connection.

station and *parking structure*. In order to decide between the two readings of the reply, then, I must suppose that the speaker wants to help the other person, and therefore is indicating a person where she can legally park, not a place where she can get gas. What matters here is that the *end* of the speaker is to help the other person, and that her words are a *means* to do so, and therefore they must fit those ends.

If that is so, then an important consequence is that the teleological structure of the action may not be clear to the speaker at the moment of the utterance, so that her intention in uttering a sentence may also be indeterminate. It may not *remain* indeterminate, however, since further interaction among the participants in the conversation can refine this teleological structure and make it clearer for those involved. Consider this example from Clark (1996, p. 213ff).[63] Suppose Jack notices that Kate is tired and then says to her, "Sit here". Is that an order? A request? An advice? Perhaps Jack himself isn't totally sure. He may think that, on the one hand, since he is her boss, that may be an order, and, on the other hand, since he is also concerned about her, and does not want to be bossy, it may be an advice. If Kate replies "What a good idea!", thus signaling her construal of the utterance as an advice, and Jack does not correct her, then he is *accepting* her construal, so that they have *together* construed the utterance as an advice.

Assuming this analysis is right, how does it apply to the truth conditions of a person's utterances? We saw during the analysis of the John Locke Lectures that Davidson then considered *speaking the truth* to be a special act, in that it had no *illouctionary force* in the sense defined in this essay, i.e. it incurred in no changes to the commitments or entitlements of the participants in the conversation and also required no uptake in order to be performed. From this, Davidson then concluded that speaking the truth was essentially *disengaged*, and therefore that the conditions for its performance could be specified in abstraction from the concrete situations in which the speech acts took place, thus rendering it ripe for systematic study. Here, I want to question Davidson's inference. Even supposing that to speak the truth is not an illocutionary act (in the sense that it has no illocutionary force attached to it), it does not follow that it is *disengaged* in the way Davidson envisions.

Consider the following example, adapted from Travis (2011, p. 188). Suppose that I am going to have dinner at Sid and Pia's place. Curious about the menu, I ask if there is going to be meat, to which Sid replies "no". Upon arrival, I discover that Sid is going to serve kidney. As I question him about this, he says that he did not lie before, since kidney is not a type of meat, but a type of *offal*. Did Sid lie to me or not? It depends. If Sid knew I was a vegan, and intended to make a cruel joke at my expense, then one can say that he lied, for kidney could be considered as meat for the purposes of the conversation. If, however, Sid

[63] McDonald (2021b, §III.1) uses a similar example to make a similar point, i.e. that the illocutionary force of an utterance is determined by the two participants in a conversations.

knew I was pedantic about certain distinctions between animal parts, he spoke the truth, for then kidney *would not* count as meat for the purposes of the conversation. Notice that whether or not Sid spoke the truth depends on the *purpose* of my question. Did I want to know if I could eat whatever it was he was cooking? Or did I want to make some fine distinctions between animal parts? As Austin puts it:

> It is essential to realize that 'true' and 'false', like 'free' and 'unfree', do not stand for anything simple at all; but only for a general dimension of being a right or proper thing to say as opposed to a wrong thing, in these circumstances, to this audience, for these purposes and with these intentions. (Austin 1962, p. 144)

In other words, against the Davidson from the John Locke Lectures, to speak the truth is not something disengaged, as determining whether or not someone spoke the truth is to *assess* whether or not a person's utterance properly describe the facts *given the purposes at hand*, that is, it is to judge whether the means fit the end. Notice that, here, we find the same phenomenon described by Clark above, that the participants must jointly decide if the person's utterances were adequate to the task or not. Consider, for example, the situation with Sid as above. Perhaps I am a vegan and *also* someone who cares for fine distinctions between animal parts, so that, even if my purpose in questioning Sid was to avoid eating animal parts, I decide that I formulated my question poorly, so that, after Sid explains himself, we come to the conclusion that he was not lying, after all. Or perhaps I am someone who cares about fine distinctions between animal parts, but Sid was unaware of this and wanted to play a joke on me, so that, after the discussion, we come to the conclusion that he did not speak the truth. Whether or not Sid spoke the truth depends then not only on how the word is, but on how his words may reasonably be taken to express how the word is, and this depends on how those involved in the conversation may jointly construe his words. That is why Travis can conclude (note the emphasis):

> Truth is *thus* the name of a general dimension of assessment of the suitability of words both for the specific situations they are used to speak of, *and* for agents who, in the circumstances of their speaking, are to take them in one way or another. (Travis 2011, p. 197)

That is, the words must be suitable also to the *agents*, so that they may reasonably construe them as adequate for describing the situation at hand. Hence, to decide upon the truth conditions of an utterance, one must be able to also discern the *purpose* and *intention* behind the utterance, that is, the *teleological unity* of the speech *act*.

Translated, then, as a practical problem, Foster's Problem can be thought of as how do we choose between different construals for the same utterance. We do so by rationalizing

the person's speech act as an action, that is, by placing it in rational pattern of behavior, by seeing how this particular act relates to the purposes and intentions at hand. This is not a fixed matter, but is open to negotiation and refinement, so that all the participants in the conversation are involved in determining the meaning of what is said. There is, thus, nothing that *is* said independently of the conversation.[64]

3.6 Conclusion

We started with a question. Did Davidson spouse a theory according to which a theory of meaning should describe the stable and systematic features of a language? Or did Davidson spouse a theory according to which a theory of meaning is something knowledge of which suffices for interpretation? In this essay, we saw how Davidson started by thinking along the lines suggested by the first question, but eventually incorporated also the line suggested in the latter question. The result, I argued, was a tension between the two demands he put on semantic theories: on the one hand, they must be the aim of the *theoretician* in describing the semantic competence of speakers, on the other hand, they must be the tools of the *interpreters* when trying to understand a speaker.

The best way to solve this tension, I argued, was to give up the idea that a theory of truth could serve as a theoretical semantics for a language. Speakers do attribute truth conditions to each other's utterances, but this attribution is context-bound and non-systematic, since it essentially depends on two dimensions, namely, the rationality of the utterance when considered against its teleological context, and the negotiation between the parts as to what this teleological context actually is. There is, therefore, no *theory* of truth for a language.[65]

Once we give the idea of a theory of truth for a language, things start to fall into place. We see that participants in a conversation come equipped with some background assumptions—the prior theory—, which they may adjust on the fly depending on their interactions with each other, thus producing passing theories. The production of passing theories are constrained by three factors: (i) the overall assumption of rationality, i.e. the constitutive ideal of rationality against which we interpret one another, (ii) the background assumptions we bring to bear, and (iii) our linguistic competence, in the form both of our association of words with bodies of information and of our syntactic competence. By employing these constraints and our *sense of occasion*, we may, together, arrive at truth—and

[64] This conclusion may call into question the notion of literal meaning itself, if that is understood as a kind of cross-contextual feature of an utterance. For analysis of Austin that support this idea, cf. Crary (2002) and Travis (2011). Hansen (2012) is a reply to Crary that tries to preserve the idea of literal meaning, but I think a more nuanced reading of Austin shows that things are not so simple; unfortunately, I do not have the space here to develop this reading of Austin. Cf. Sbisà (2015) for a start.

[65] In this, then, I believe Hacking (1986) was right.

whatever other goals we may also have in communicating. This is, I believe, the real sense in which language is *social*: not in that it involves conventional straight-jackets, but in that it involves a collective endeavor to arrive at a *common* goal; not in what it *constraints*, but in what it *allows*.

Conclusion

After finishing the three chapters that compose this thesis, the reader may be wondering what all those haphazard remarks have in common. Does it add up to a coherent story, after all? Well, I believe it does, and here I would like to present what seems to me to be the bigger picture.

We start life with an innate linguistic endowment, which, following Chomsky, we may call *Universal Grammar*. This innate endowment does not consist on surface level features (more akin to Greenberg universals—cf. Hornstein (2019) for some comments on the difference), but on certain principles and parameters that determine the form of our adult linguistic competence. The adult, stable form of this innate linguistic endowment is called an *I-Language*; as stressed by Collins (2010, p. 47), this is not an "idiolect", at least not if an idiolect is thought of as a mind-independent entity, specifiable in an externalist fashion. Instead, it is merely the state of a subject's faculty, the faculty of language, and we may hope to one day describe its implementation in the biology of the individual. In the meantime, we employ our theoretical idiom to describe the structural relations among the states of this faculty, knowing full well that this is a temporary measure, a placeholder for the fuller biolinguistic description—similarly to how the dispositional idiom is a promissory note for both Quine and Davidson, as analyzed in Chapters 1 and 3, respectively.

One important landmark in the individual's developmental path toward the maturation of her I-Language is the acquisition of predicate frame structures. These predicate frame structures form the basis for the assignment of thematic roles in a sentence, and hence are a tool for the conceptual organization of a thought. Unsurprisingly, it is the appearance of such frames in the homesign of children that is one of the hallmarks of a linguistic system (Goldin-Meadow 2003, chap. 12); once they appear, they lead to a complete reorganization across the child's language, with gestures losing their iconic status as they are now recruited into this larger system. This is not hyperbolic; I would claim that this reorganization of thought driven by the process of language acquisition is one of the main landmarks in the cognitive development of an individual and of our species. It is the main contribution of language for thought, especially since its recursivity allows for the creation of concepts of

ever increasing complexity.[1]

The main point of Chapter 1 is that this faculty, this I-Language, is all that we can hope to study scientifically. Quine was right when he proclaimed the "baselessness of intentional idioms and the emptiness of a science of intention" (Quine 1960/2013, p. 202). As I hope I have cogently argued at least for the case of truth conditions, intentional idioms are not fit for science, because they depend on *our interpretation* of the individual to which we are ascribing a representation. As Quine says, what is involved in interpretation is an "evaluation, relative to special purposes, of an essentially dramatic act" (Quine 1960/2013, p. 200). In such an act, "we do not generally know how much reality to hold constant. Quandaries arise" (ibid.). Being thus relative to the special purposes of the occasion, whether an interpretation is cogent or not can change depending on such purposes. There is thus no science of interpretation, since science deals precisely with what is held constant across contexts, with what is stable. In order to interpret, we need a *sense of occasion*, a sensitivity to the particular demands of a situation that is simply outside the domain of proper science.

If the semantic notions are then thought of in an intentional way—say, as involving the concepts of truth and reference—, then it follows that there can be no science of semantics. As Pietroski (2018) reminds us, however, this need not be the only way of conceiving of semantics. We can think of semantics as one of the capacities that enables our activity of interpretation, by sufficiently constraining it so that reasonable interpretations can be produced. As per the reasoning in the preceding paragraph, this does not mean that what is *reasonable* is context-insensitive, on the contrary. But it does mean that what is reasonable in a context is in part constrained by the semantics of the utterance. Here, again, we see the importance of the argument structure of a sentence, since that is one of the main constraints of interpretation, namely that it respects this structure. This is what allows for the process of syntactic bootstrapping, initially described by Landau and Gleitman (1985) to understand how congenitally blind children manage to interpret and use sight-related words, and which I claim is a ubiquitous process in utterance interpretation.[2]

So stripped down, then, semantics becomes much more similar to the original proposal by Katz and Fodor (1963/1964) about how to integrate semantics into a generative framework. As I envision it, it has three main components. One is the study of whatever constraints are associated with argument structure, such as the study of thematic roles. Another is the study of whatever constraints are associated with particular lexical items, perhaps as attenuated forms of semantic markers, as in Katz and Fodor (1963/1964),[3] or

[1] Cf. Reboul (2017) for a speculative evolutionary narrative that postulates this reorganization as the main evolutionary advantage promoted by the emergence of language.

[2] Cf. also Borer (2004) for the importance of this process.

[3] Lewis complained that "we can know the Markerese translation of an English sentence without knowing the first thing about the meaning of the English sentence: namely, the conditions under which it would be

perhaps something more in the lines of the recent proposal by Harris (2020a). Finally, it is also the study of how the lexicalization process drives our conceptual organization, by studying how it provides instructions for the building of concepts (Pietroski 2018). If the label was not already taken, one could say that this is a form of *semantic minimalism*, since, in this proposal, the semantic features of an utterance are truly minimal, so that most of the interpretive work is done by pragmatic processes. Perhaps "bare-bones semantics" or "skeletal semantics" (alluding to Borer's "exo-skeletal" approach) could be appropriate labels.

Importantly, this serves to distinguish my proposal from the proposal of, e.g., Jackendoff (2003). For Jackendoff, the meaning of a word is the concept associated with it (the word is an interface between different cognitive modules, including the phonological, the syntactic, and the conceptual). Hence, for Jackendoff, semantics should study not only the linguistic constraints introduced by the grammar of a sentence, but also the conceptual organization of the lexicon. For reasons explored in Chapter 3, I do not think this is feasible. As Davidson argues, there is no sharp distinction between our knowledge of the world and our knowledge of words, in the sense that there is no stable core to the body of information associated with a word that could serve as *the* meaning of the word. Rather, in hearing an utterance, we receive an instruction to retrieve the *entire* body of information, and it is only our sense of the occasion, our sensitivity to the particular demands of the context, that allows us to select whatever is relevant to evaluate the utterance at hand. That is not to say that concepts cannot have interesting structure, even in the absence of linguistic capacities. But this is for semantics to study. This is for cognitive science more generally to study, with special emphasis on developmental psychology and animal studies.

Chapters 2 and 3, then, are, in a sense, about the perils of ignoring the Quinean strictures proposed in Chapter 1. We see there how Lewis and Davidson face insurmountable obstacles in trying to give a systematic treatment of truth conditions. In the case of Lewis, the difficulty is compounded by his insistence on studying language abstractly, as if it were a mathematical object, instead of focusing on the linguistic competence of the individuals—i.e. their internalized grammars. In the case of Davidson, the difficulty is partly mitigated by his acute sensibility—as displayed in "A Nice Derangement of Epitaphs"—to the dynamics of interpretation, though, as we saw, this same sensibility serves to undercut the rationale for a systematic science of truth conditions. These chapters serve as a kind of cautionary tale for the confusion that arises when one ignores the Quinean strictures.

The overall lesson is not wholly negative, though. For both Lewis and Davidson also contributed invaluable insights into the process of interpretation. Once one reframes their contributions as contributions not to semantics, but to pragmatics, everything begins to fall

true" (Lewis 1970/1983a, p. 190). From my perspective, this is exactly right, but it is a *virtue* of "Markerese" approaches, not a vice.

into place. In particular, this allows one to abandon Lewis's code model for a language, and thus to give full weight to his idea of communication as a *joint coordination project*, that is, as a *joint activity* undertaken by the participants to achieve certain goals (Clark 1996). As Stalnaker saw, this means that there is a body of information that is accepted for the purposes of communication, the common ground, and that we can consider the effects of various linguistic acts to this body. It also means, as Lewis saw, that the conversation dynamics can evolve inside certain boundaries established by linguistic conventions, and, moreover, that some of these conventions have as effects the change in normative commitments of the participants, as Austin originally suggested (deontic scorekeeping). Finally, it means that, as Davidson saw, the meaning of an utterance is many times established *after* the utterance was made, so that participants often engage in negotiations as to what meaning should be attributed to the utterance, negotiations that are a species of *joint construals* in the terminology of Clark (1996). This is the sense in which communication is truly *social*, and not merely a series of autonomous actions.

By insisting, then, on a sharp separation between semantics and pragmatics, and keeping the intentional to the side of pragmatics, we can better appreciate, on the one hand, what is peculiar to our biological endowment and what is peculiar to our social interactions. It also allows us to better recognize how flexible a tool language is, how it enables novel and creative uses of our own cognitive capacities. The fact that the argument structure of a sentence forces a certain thematic interpretation on the concepts involved also means that we can use this structure to build new concepts, by coercing old ones into new positions. The fact that linguistic communication is so underdetermined by linguistic conventions means that there is more room for social interactions to fix this underdetermination, more room for collaborative processes even when it comes to establishing what an utterance meant. These dimensions support each other. Semantics cannot determine univocally what is said, because otherwise language would be too rigid to serve its social function; but, conversely, the social function of language also calls for something stable to guide the interpretation of the participants, for a game with no rules is no game at all.

I emphasize this last point because there are common complaints leveled against Chomsky and Travis that, I believe, rest on misunderstanding this very point. Against Chomsky, it is often complained that he ignores the diversity of human languages and how social forces shape our linguistic lives. Against Travis, it is often complained that he ignores the conventional forces at work in our communication, thus making communicative success a miracle. We are now in a position to see why these complaints are misguided. Chomsky does not ignore the diversity of human languages, nor does he ignore how social forces shape our linguistic lives (indeed, he is very attuned to how social and political forces shape our lives, linguistically or otherwise). True, he emphasizes the universal aspect beneath the

diversity, but that is because he is insisting on investigating the stable core of language that allows for this diversity in the first place; this core is the only feature of language that can be investigated scientifically, anyway. That is not say other aspects are irrelevant, just that they are not as amenable to systematic treatment as that stable component is. Similarly, Travis emphasizes the reverse side of this, namely that pragmatic factors are precisely not amenable to systematic treatment. Indeed, the lesson from Travis's elaborate examples is that so much of our communicative interactions depends on our sensibility to the forces at work in a particular context. It is this attunement to each other and to the context that allows for communication to succeed without miracles. True, we do not have the security of conventions when interpreting one another. But we also do not need it. We have each other.

References

Alexiadou, Artemis and Terje Lohndal. "From the Origins of Government and Binding to the Current State of Minimalism". In: *A Companion to Chomsky*. Ed. by Nicholas Allot, Terje Lohdal, and Georges Rey. Hoboken, New Jersey: Wiley Blackwell, 2021, pp. 25–51.

Allott, Nicholas and Neil Smith. "Chomsky and Fodor on Modularity". In: *A Companion to Chomsky*. Ed. by Nicholas Allot, Terje Lohdal, and Georges Rey. Hoboken, New Jersey: Wiley Blackwell, 2021, pp. 529–543.

Anscombe, Gertrude Elizabeth Margaret. *Intention*. 2nd ed. Cambridge, Massachusetts: Harvard University Press, 1963.

Antonsen, Paal. "Scorekeeping". In: *Analysis* 78 (2018), pp. 589–595.

Antony, Louise. "Meaning and Semantic Knowledge". In: *Proceedings of the Aristotelian Society* sup. 71 (1997), pp. 177–209.

Austin, John Langshaw. *How To Do Things With Words*. Oxford: Oxford University Press, 1962.

Bar-On, Dorit. "On the Possibility of a Solitary Language". In: *Noûs* 26.1 (1992), pp. 27–45.

— "Semantic Verificationism, Linguistic Behaviorism, and Translation". In: *Philosophical Studies* 66.3 (1992), pp. 235–359.

Begby, Endre. "Deranging the Mental Lexicon". In: *Inquiry* 59.1 (2016), pp. 33–55.

— "Language from the Ground Up: A Study of Homesign Communication". In: *Erkenntnis* 82 (2017), pp. 693–714.

Bloomfield, Leonard. *Language*. London: George Allen & Unwin, 1933/1961.

Boeckx, Cedric. *Language in Cognition: Uncovering Mental Structures and the Rules Behind Them*. Oxford: Wiley-Blackwell, 2010.

Borer, Hagit. "Exo-skeletal vs. endo-skeletal explanations: syntactic projections and the lexicon". In: *The Nature of Explanation in Linguistic Theory*. Ed. by John Moore and Maria Polinsky. Stanford, CA: Center for the Study of Language and Information, 2003, pp. 31–67.

Borer, Hagit. "The Grammar Machine". In: *The Unaccusativity Puzzle: Explorations of the Syntax-Lexicon Interface*. Ed. by Artemis Alexiadou, Elena Anagnostopoulou, and Martin Everaert. Oxford: Oxford University Press, 2004, pp. 288–331.

— *Structuring Sense: In Name Only*. Vol. 1. Oxford: Oxford University Press, 2005.

— *Structuring Sense: The Normal Course of Events*. Vol. 2. Oxford: Oxford University Press, 2005.

— *Structuring Sense: Taking Form*. Vol. 3. Oxford: Oxford University Press, 2013.

Brandom, Robert Boyce. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Massachusetts: Harvard University Press, 1994.

Brown, Jessica and Herman Cappelen, eds. *Assertion: New Philosophical Essays*. Oxford: Oxford University Press, 2011.

Burge, Tyler. "Demonstrative Constructions, Reference, and Truth". In: *Journal of Philosophy* 71 (1974), pp. 205–223.

Burgess, John P. "Tarski's Tort". In: *Mathematics, Models, and Modality: Selected Philosophical Essays*. Cambridge: Cambridge University Press, 2008, pp. 149–168.

Camp, Elisabeth. "Conventions' Revenge: Davidson, Derangement, and Dormativity". In: *Inquiry* 59.1 (2016), pp. 113–138.

Cappelen, Herman and Ernest Lepore. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Malden, MA: Blackwell Publishing, 2005.

Carnap, Rudolf. "The methodological character of theoretical concepts". In: *Minnesota Studies in the Philosophy of Science, Volume 1: The foundations of science and the concepts of psychology and psychoanalysis*. Ed. by Herbert Feigl and Michael Scriven. Minneapolis: University of Minnesota Press, 1956, pp. 38–76.

— *An Introduction to the Philosophy of Science*. Ed. by Martin Gardner. New York: Dover, 1974/1995.

Chomsky, Noam. "Three Models for the Description of Language". In: *IRE Transactions on Information Theory* 2 (1956), pp. 113–123.

— *Syntactic Structures*. The Hague: Mouton, 1957.

— "Review of B. F. Skinner's *Verbal Behavior*". In: *Language* 35.1 (1959), pp. 26–58.

— *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The MIT Press, 1965.

— *Reflections on Language*. New York: Pantheon Books, 1975.

— *Lectures on Government and Binding: the Pisa Lectures*. 7th ed. Berlin: Mouton de Gruyter, 1981/1993.

— *Knowledge of Language: Its Nature, Origins, and Use*. New York: Praeger, 1986.

— *The Minimalist Program*. Cambridge, Massachusetts: The MIT Press, 1995/2015.

— *On Nature and Language*. Cambridge: Cambridge University Press, 2002.

— "Reply to Gopnik". In: *Chomsky and his Critics*. Ed. by Louise M. Antony and Norbert Hornstein. Oxford: Blackwell, 2003, pp. 316–325.
— "Reply to Millikan". In: *Chomsky and his Critics*. Ed. by Louise M. Antony and Norbert Hornstein. Oxford: Blackwell, 2003, pp. 308–315.
— "Two Notions of Modularity". In: *On Concepts, Modules, and Language*. Ed. by Roberto G. de Almeida and Lila R. Gleitman. Oxford: Oxford University Press, 2018, pp. 25–40.
Churchland, Paul Montgomery. "On the nature of theories: a neurocomputational perspective". In: *Minnesota Studies in the Philosophy of Science* 14 (1989). Ed. by C. W. Savage, pp. 59–101.
— "Cognitive activity in artificial neural networks". In: *An Invitation to Cognitive Science, Volume 3: Thinking*. Ed. by Daniel N. Osherson and Edward E. Smith. Cambridge, Massachusetts: MIT Press, 1990, pp. 199–227.
Clapp, Lenny. "Stalnaker on the Essential Effect of Assertion". In: *The Oxford Handbook of Assertion*. Ed. by Sanford Goldberg. Oxford: Oxford University Press, 2020, pp. 3–38.
Clark, Herbert H. *Using Language*. Cambridge: Cambridge University Press, 1996.
Collins, John. "On the Very Idea of a Science Forming Faculty". In: *Dialectica* 56.2 (2002), pp. 125–151.
— "Faculty Disputes". In: *Mind and Language* 19.5 (2004), pp. 503–533.
— "Meta-Scientific Eliminativism: A Reconsideration of Chomsky's Review of Skinner's *Verbal Behavior*". In: *The British Journal for the Philosophy of Science* 58.4 (2007), pp. 625–658.
— *Chomsky: A Guide for the Perplexed*. London: Continuum, 2008.
— "Naturalism in the Philosophy of Language; or Why There Is No Such Thing as Language". In: *New Waves in Philosophy of Language*. Ed. by Sarah Sawyer. New York: Palgrave Macmillan, 2010, pp. 41–59.
— "Representations without Representa: Content and Illusion in Linguistic Theory". In: *Semantics and Beyond: Philosophical and Linguistic Inquiries*. Ed. by Piotr Stalmaszczyk. Berlin: De Gruyter, 2014, pp. 27–64.
— "Perceiving Language: Issues between Travis and Chomsky". In: *The Philosophy of Charles Travis: Language, Thought, and Perception*. Ed. by John Collins and Tamara Dobler. Oxford: Oxford University Press, 2018, pp. 155–180.
— *Linguistic Pragmatism and Weather Reporting*. Oxford: Oxford University Press, 2020.
Collins, John and Georges Rey. "Chomsky and Intentionality". In: *A Companion to Chomsky*. Ed. by Nicholas Allot, Terje Lohdal, and Georges Rey. Hoboken, New Jersey: Wiley Blackwell, 2021, pp. 488–502.
Crary, Alice. "The Happy Truth: J. L. Austin's *How To Do Things With Words*". In: *Inquiry* 45.1 (2002), pp. 59–80.

Davidson, Donald. "Actions, Reasons, and Causes". In: *Essays on Actions and Events*. 2nd ed. Oxford: Clarendon Press, 1963/2001, pp. 3–19.
— "Theories of Meaning and Learnable Languages". In: *Inquiries Into Truth and Interpretation*. 2nd ed. New York: Oxford University Press, 1965/2001, pp. 3–15.
— "Causal Relations". In: *Essays on Actions and Events*. 2nd ed. Oxford: Clarendon Press, 1967/2001, pp. 63–81.
— "True to the Facts". In: *Inquiries into Truth and Interpretation*. 2nd ed. New York: Oxford University Press, 1969/2001, pp. 43–54.
— "Semantics for Natural Languages". In: *Inquiries into Truth and Interpretation*. 2nd ed. New York: Oxford University Press, 1970/2001, pp. 55–64.
— *The Structure of Truth: The 1970 John Locke Lectures*. Ed. by Cameron Domenico Kirk-Giannini and Ernest Lepore. Oxford: Oxford University Press, 1970/2020.
— "Freedom to Act". In: *Essays on Actions and Events*. 2nd ed. Oxford: Clarendon Press, 1973/2001, pp. 149–162.
— "Radical Interpretation". In: *Inquiries into Truth and Interpretation*. 2nd ed. New York: Oxford University Press, 1973/2001, pp. 125–139.
— "Belief and the Basis of Meaning". In: *Inquiries into Truth and Interpretation*. 2nd ed. New York: Oxford University Press, 1974/2001, pp. 141–154.
— "Thought and Talk". In: *Inquiries into Truth and Interpretation*. 2nd ed. New York: Oxford University Press, 1975/2001, pp. 155–170.
— "Reply to Foster". In: *Inquiries into Truth and Interpretation*. 2nd ed. New York: Oxford University Press, 1976/2001, pp. 171–179.
— "What Metaphors Mean". In: *Inquiries into Truth and Interpretation*. 2nd ed. New York: Oxford University Press, 1978/2001, pp. 245–264.
— "Mental Events". In: *Essays on Actions and Events*. 2nd ed. Oxford: Clarendon Press, 1980/2001, pp. 207–225.
— "A Unified Theory of Thought, Meaning, and Action". In: *Problems of Rationality*. Oxford: Clarendon Press, 1980/2004, pp. 101–116.
— "Communication and Convention". In: *Inquiries into Truth and Interpretation*. 2nd ed. New York: Oxford University Press, 1984/2001, pp. 265–280.
— "A Nice Derangement of Epitaphs". In: *Truth, Language, and History*. New York: Oxford University Press, 1986/2005, pp. 89–107.
— "Problems in the Explanation of Action". In: *Problems of Rationality*. Oxford: Clarendon Press, 1987/2004, pp. 151–166.
— "What is Present to the Mind?" In: *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press, 1989/2001, pp. 53–67.

— "Representation and Interpretation". In: *Problems of Rationality*. Oxford: Clarendon Press, 1990/2004, pp. 117–134.
— "Reply to Ralf Stoecker". In: *Reflectin Davidson: Donald Davidson Responding to an International Forum of Philosophers*. Ed. by Ralf Stoecker. Berlin: De Gruyter, 1993, pp. 287–290.
— "Could There Be a Science of Rationality?" In: *Problems of Rationality*. Oxford: Clarendon Press, 1995/2004, pp. 87–99.
Davies, Alex. "Communicating by doing something else". In: *The Philosophy of Charles Travis: Language, Thought, and Perception*. Ed. by John Collins and Tamara Dobler. Oxford: Oxford University Press, 2018, pp. 113–134.
Davies, Martin. "Meaning, Structure and Understanding". In: *Synthese* 48 (1981), pp. 135–162.
— "Tacit Knowledge and Semantic Theory: Can a Five Per Cent Difference Matter?" In: *Mind* 96.384 (1987), pp. 441–462.
Drobňak, Matej. "Quine on Shared Language and Linguistic Communities". In: *Philosophia* 46 (2018), pp. 83–99.
Dummett, Michael. "A Nice Derangement of Epitaphs: Some Comments on Davidson and Hacking". In: *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Ed. by Ernest Lepore. Cambridge: Blackwell, 1986, pp. 459–476.
Ebbs, Gary. "Review of *The Pursuit of Truth*". In: *The Philosophical Review* 103.3 (1994), pp. 535–541.
Evans, Gareth. "Semantic Theory and Tacit Knowledge". In: *Collected Papers*. Ed. by Antonia Phillips. Oxford: Clarendon Press, 1981/1985, pp. 322–342.
Feldman, Heidi, Susan Goldin-Meadow, and Lila Gleitman. "Beyond Herodotus: The creation of a language by linguistically deprived deaf children". In: *Action, Symbol, and Gesture: The Emergence of Language*. Ed. by Andrew Lock. New York: Academic Press, 1978, pp. 351–414.
Fisher, Tyrus. "Quine's Behaviorism and Linguistic Meaning: Why Quine's Behaviorism is not Illicit". In: *Philosophia* 39 (2011), pp. 51–59.
Føllesdal, Dagfinn. "In what sense is language public?" In: *On Quine*. Ed. by Paolo Leonardi and Marco Santambrogio. Cambridge: Cambridge University Press, 1995, pp. 53–67.
Foster, John. "Meaning and Truth Theory". In: *Truth and Meaning: Essays in Semantics*. Ed. by Gareth Evans and John Henry McDowell. Oxford: Clarendon Press, 1976, pp. 1–32.
Fricker, Elizabeth. "Semantic Structure and Speakers' Understanding". In: *Proceedings of the Aristotelian Society* 83 (1982-1983), pp. 49–66.

Fuller, Tim. "Is scientific theory change similar to early cognitive development? Gopnik on science and childhood". In: *Philosophical Psychology* 26.1 (2013), pp. 109–128.

George, Alexander. "Review discussion: *Knowledge of Language: Its Nature, Origin, and Use*, by Noam Chomsky". In: *Mind and Language* 2.2 (1987), pp. 155–164.

— "Has Dummett Oversalted His Frege? Remarks on the Conveyability of Thought". In: *Language, Thought, and Logic: Essays in Honour of Michael Dummett*. Ed. by Richard Kimberly Heck. (Originally edited under the name "Richard G. Heck, Jr". Oxford: Oxford University Press, 1997, pp. 35–69.

— "Quine and Observation". In: *Knowledge, Language and Logic: Questions for Quine*. Ed. by Alex Orenstein and Petr Kotatko. Dordrecht: Kluwer Academic Publishers, 2000, pp. 21–45.

Geurts, Bart. "Communication as commitment sharing: speech acts, implicatures, common ground". In: *Theoretical Linguistics* 45.1-2 (2019), pp. 1–30.

Gleitman, Lila R. *Sentence First, Arguments Afterward: Essays in Language and Learning*. Ed. by Jeffrey Lidz. With a forew. by Noam Chomsky. Oxford: Oxford University Press, 2020.

Glüer, Kathrin. *Donald Davidson: A Short Introduction*. Oxford: Oxford University Press, 2011.

Goldberg, Sanford C. *Assertion: On the Philosophical Significance of Assertoric Speech*. Oxford: Oxford University Press, 2015.

— ed. *The Oxford Handbook of Assertion*. Oxford: Oxford University Press, 2020.

Goldin-Meadow, Susan. *The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us About How All Children Learn Language*. New York: Psychology Press, 2003.

Gopnik, Alison. "The Theory Theory as an Alternative to the Innateness Hypothesis". In: *Chomsky and his Critics*. Ed. by Louise M. Antony and Norbert Hornstein. Oxford: Blackwell, 2003, pp. 238–254.

Grant, Cosmo. "Mistakes About Conventions and Meanings". In: *Topoi* (2019), pp. 1–15.

Green, Mitchell S. "On the Autonomy of Linguistic Meaning". In: *Mind* 106.422 (1997), pp. 217–243.

Greenberg, Gabriel. "The Iconic-Symbolic Spectrum". In: (). Manuscript.

Gregory, Paul A. *Quine's Naturalism: Language, Theory and the Knowing Subject*. London: Continuum, 2008.

Gustafsson, Martin. "Systematic Meaning and Linguistic Diversity: The Place of Meaning-Theories in Davidson's Later Philosophy". In: *Inquiry* 41.4 (1998), pp. 435–453.

Hacking, Ian. "The Parody of Conversation". In: *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Ed. by Ernest Lepore. Cambridge: Blackwell, 1986, pp. 447–458.

— "Natural Kinds". In: *Perspectives on Quine*. Ed. by Robert Barrett and Roger Gibson. Oxford: Blackwell Publishers, 1990, pp. 129–141.

Hansen, Nat. "J. L. Austin and Literal Meaning". In: *European Journal of Philosophy* 22.4 (2012), pp. 617–632.

Harris, Daniel W. "Semantics without semantic content". In: *Mind and Language* (2020). Early view, pp. 1–25.

— "We talk to people, not contexts". In: *Philosophical Studies* 177 (2020), pp. 2713–2733.

Harris, Daniel W., Daniel Fogal, and Matt Moss. "Speech Acts: The Contemporary Landscape". In: *New Work on Speech Acts*. Ed. by Daniel W. Harris, Daniel Fogal, and Matt Moss. Oxford: Oxford University Press, 2018, pp. 1–39.

Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" In: *Science* 298.5598 (2002), pp. 1569–1579.

Hawthorne, John. "A note on 'Languages and Language'". In: *Australasian Journal of Philosophy* 68 (1 1990), pp. 116–118.

Heck, Richard Kimberly. "Semantics and Context-Dependence: Towards a Strawsonian Account". In: *Metasemantics: New Essays on the Foundations of Meaning*. Ed. by Alexis Burgess and Brett Sherman. Originally published under the name "Richard G. Heck, Jr." Oxford: Oxford University Press, 2014, pp. 327–364.

Higginbotham, James. "Contexts, Models, and Meanings: A Note on the Data of Semantics". In: *Mental Representations: The Interface Between Language and Reality*. Ed. by Ruth Margaret Kempson. Cambridge: Cambridge University Press, 1988, pp. 29–48.

— "Competence with Demonstratives". In: *Philosophical Perspectives* 16 (2002), pp. 1–16.

Hoff, Erika. "How social context support and shape language development". In: *Developmental Review* 26 (2006), pp. 55–88.

— *Language Development*. 5th ed. Belmont, CA: Wadsworth, 2013.

Hornsby, Jennifer and Rae Langton. "Free Speech and Illocution". In: *Legal Theory* 4 (1998), pp. 21–37.

Hornstein, Norbert. "The Stupendous Success of the Minimalist Program". In: *Current Approaches to Syntax: A Comparative Handbook*. Ed. by András Kertész, Edith Moravcsik, and Csilla Rákosi. Berlin: De Gruyter Mouton, 2019, pp. 187–214.

Hylton, Peter. *Quine*. London: Routledge, 2007.

Jackendoff, Ray. *Semantics and Cognition*. Cambridge, Massachusetts: The MIT Press, 1983.

— *Foundations of Language: brain, Meaning, Grammmar, Evolution*. Oxford: Oxford University Press, 2003.

Jankovic, Marija and Greg Ray. "Meaning, Publicity and Knowledge". In: *ProtoSociology* 34 (2017), pp. 98–115.

Jeffrey, Richard Carl. *The Logic of Decision*. 2nd ed. Chicago: University of Chicago Press, 1990.

Katz, Jerrold J. and Jerry A. Fodor. "The Structure of a Semantic Theory". In: *The Structure of Language: Readings in the Philosophy of Language*. Ed. by Jeery A. Fodor and Jerrold J. Katz. New Jersey: Prentice-Hall, 1963/1964, pp. 479–518.

Katz, Jerrold Jacob. *Language and Other Abstract Objects*. Oxford: Basil Blackwell, 1981.

Kemp, Gary. *Quine: A Guide for the Perplexed*. New York: Continuum, 2006.

— *Quine versus Davidson: Truth, Reference, and Meaning*. Oxford: Oxford University Press, 2012.

— "Quine, Publicity, and Pre-Established Harmony". In: *ProtoSociology* 34 (2017), pp. 59–72.

— "Quine Peregrinating: Norms, Dispositions, and Analyticity". In: *From Rules to Meanings: New Essays on Inferentialism*. Ed. by Ondřej Beran, Voltěch Kolman, and Ladislav Koreň. New York: Routledge, 2018, pp. 81–96.

— "Observation Sentences Revisited". In: *Mind* (2021). Advance Article.

Lance, Mark Norris and Rebecca Kukla. *'Yo!' and 'Lo!': The Pragmatic Topography of the Space of Reasons*. Cambridge, Massachusetts: Harvard University Press, 2009.

Landau, Barbara and Lila Gleitman. *Language and Experience: Evidence from the Blind Child*. Cambridge, Massachusetts: Harvard University Press, 1985.

Lashley, Karl Spencer. "The Problem of Serial Order in Behavior". In: *Cerebral Mechanisms in Behavior*. Ed. by L. A. Jeffress. New York: Wiley, 1951, pp. 112–131.

Lenneberg, Eric Heinz. "The Capacity for Language Acquisition". In: *The Structure of Language: Readings in the Philosophy of Language*. Ed. by Jerry A. Fodor and Jerrold Jacob Katz. New Jersey: Prentice-Hall, 1964, pp. 579–603.

— *Biological Foundations of Language*. New York: Wiley, 1967.

Lepore, Ernest and Kirk Ludwig. *Donald Davidson: Meaning, Truth, Language, and Reality*. New York: Oxford University Press, 2005.

Lepore, Ernest and Matthew Stone. "David Lewis on Convention". In: *A Companion to David Lewis*. Ed. by Barry Loewer and Jonathan Schaffer. Malden, MA: Wiley-Blackwell, 2015, pp. 315–327.

— *Imagination and Convention: : Distinguishing Grammar and Inference in Language*. Oxford: Oxford University Press, 2015.

— "Convention before Communication". In: *Philosophical Perspectives* 31.1 (2017), pp. 245–265.

Lewis, David Kellogg. *Convention*. Cambridge, Massachusetts: Harvard University Press, 1969.

— "General Semantics". In: *Philosophical Papers*. Vol. 1. New York: Oxford University Press, 1970/1983, pp. 189–232.
— "How to Define Theoretical Terms". In: *Philosophical Papers*. Vol. 1. New York: Oxford University Press, 1970/1983, pp. 78–95.
— "Radical Interpretation". In: *Philosophical Papers*. Vol. 1. New York: Oxford University Press, 1974/1983, pp. 108–118.
— "Language and Languages". In: *Philosophical Papers*. Vol. 1. New York: Oxford University Press, 1975/1983, pp. 163–188.
— "Scorekeeping in a Language Game". In: *Philosophical Papers*. Vol. 1. New York: Oxford University Press, 1979/1983, pp. 233–249.
— "Index, Context, and Content". In: *Papers in Philosophical Logic*. Cambridge: Cambridge University Press, 1980/1998, pp. 21–44.
— "New Work for a Theory of Universals". In: *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press, 1983/2000, pp. 8–65.
— "Meaning without use: Reply to Hawthorne". In: *Papers in Ethics and Social Philosophy*. Cambridge: Cambridge University Press, 1992/2000, pp. 145–151.
Loar, Brian. *Mind and Meaning*. Cambridge: Cambridge University Press, 1981.
Machery, Edouard. *Doing without Concepts*. Oxford: Oxford University Press, 2009.
— "Replies to my critics". In: *Philosophical Studies* 149.3 (2010), pp. 429–436.
McDonald, Lucy. "Reimagining Illocutionary Force". In: *The Philosophical Quarterly* (2021). Advance Article. doi: https://doi.org/10.1093/pq/pqab065.
— "Your word against mine: the power of uptake". In: *Synthese* 199 (2021), pp. 3505–3526.
McGilvray, James. "Meanings Are Syntactically Individuated and Found in the Head". In: *Mind and Language* 13.2 (1998), pp. 225–280.
McKinnon, Rachel. *The Norms of Assertion: Truth, Lies, and Warrant*. New York: Palgrave Macmillan, 2015.
Millikan, Ruth Garrett. "Pushmi-pullyu Representations". In: *Language: A Biological Model*. Oxford: Oxford University Press, 1995/2005, pp. 166–186.
Morgan, Gary. "Review of Susan Goldin-Meadow's *The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us about How All Children Learn Language*". In: *Journal of Child Language* 32.4 (2005), pp. 925–928.
Newmeyer, Frederick J. *Linguistic Theory in America*. 2nd ed. Orlando: Academic Press, 1986.
Pietroski, Paul. *Conjoining Meanings: Semantics Without Truth Values*. Oxford: Oxford University Press, 2018.
Pietroski, Paul et al. "The Meaning of ' Most': Semantics, Numerosity, and Psychology". In: *Mind and Language* 24 (2009), pp. 554–585.

Pullum, Geoffrey Keith. "On the Mathematical Foundations of *Syntactic Structures*". In: *Journal of Logic, Language and Information* 20.3 (2011), pp. 277–296.

Quine, Willard Van Orman. *Word and Object*. 2nd ed. Cambridge, Massachusetts: MIT Press, 1960/2013.

— "Linguistics and Philosophy". In: *The Ways of Paradox and Other Essays*. Cambridge, Massachusetts: Harvard University Press, 1968/1975, pp. 56–58.

— "Epistemology Naturalized". In: *Ontological Relativity and Other Essays*. New York: Columbia University Press, 1969, pp. 69–90.

— "Natural Kinds". In: *Ontological Relativity and Other Essays*. New York: Columbia University Press, 1969, pp. 114–138.

— "Ontological Relativity". In: *Ontological Relativity and Other Essays*. New York: Columbia University Press, 1969, pp. 26–68.

— *The Roots of Reference*. La Salle, Illinois: Open Court, 1973.

— "Mind and Verbal Dispositions". In: *Mind and Language*. Ed. by Samuel Guttenplan. Oxford: Oxford University Press, 1975, pp. 83–95.

— "Empirical Content". In: *Theories and Things*. Cambridge, Massachusetts: Harvard University Press, 1981, pp. 24–30.

— "Indeterminacy of Translation Again". In: *The Journal of Philosophy* 84.1 (1987), pp. 5–10.

— "Comment on Hacking". In: *Perspectives on Quine*. Ed. by Robert Barrett and Roger Gibson. Oxford: Blackwell Publishers, 1990, pp. 142–143.

— "Three Indeterminacies". In: *Perspectives on Quine*. Ed. by Robert Barrett and Roger Gibson. Oxford: Blackwell Publishers, 1990, pp. 1–16.

— *Pursuit of Truth*. Revised. Cambridge, Massachusetts: Harvard University Press, 1992.

— "In Praise of Observation Sentences". In: *The Journal of Philosophy* 90.3 (1993), pp. 107–116.

— *From Stimulus to Science*. Cambridge, Massachusetts: Havrad University Press, 1995.

— "Preestablished Harmony". In: *Quine and His Place in History*. Ed. by Frederique Janssen-Lauret and Gary Kemp. New York: Palgrave Macmillan, 1995/2016, pp. 29–32.

— "Response to Gary Ebbs". In: *Quine and His Place in History*. Ed. by Frederique Janssen-Lauret and Gary Kemp. New York: Palgrave Macmillan, 1995/2016, pp. 33–36.

— "Progress on Two Fronts". In: *The Journal of Philosophy* 93.4 (1996), pp. 159–163.

— "Response to George". In: *Knowledge, Language and Logic: Questions for Quine*. Ed. by Alex Orenstein and Petr Kotatko. Dordrecht: Kluwer Academic Publishers, 2000, pp. 408–410.

Reboul, Anne. *Cognition and Communication in the Evolution of Language*. Oxford: Oxford University Press, 2017.

Rescorla, Michael. "Rationality as a Constitutive Ideal". In: *A Companion to Donald Davidson*. Ed. by Ernest Lepore and Kirk Ludwig. West Sussex: Wiley Blackwell, 2013, pp. 472–488.

— "Convention". In: *The Stanford Encyclopedia of Philosophy (Summer 2019 Edition)*. Ed. by Edward N. Zalta. 2019. url: https://plato.stanford.edu/archives/sum2019/entries/convention/.

Rowland, Caroline. *Understanding Child Language Acquisition*. New York: Routledge, 2014.

Rumfitt, Ian. "Review of Donald Davidson's *The Structure of Truth: The 1970 John Locke Lectures*". In: *Mind* (2021). Online First.

Saxton, Matthew. *Child Language: Acquisition and Development*. 2nd ed. London: SAGE, 2017.

Sbisà, Marina. "Speech acts in context". In: *Language and Communication* 22 (2002), pp. 421–436.

— "Uptake and conventionality in illocution". In: *Lodz Papers in Pragmatics* 5 (2009), pp. 33–52.

— "C'è del pragmatismo in J.L. Austin? Una rilettura delle proposte austiniane sul tema della verità". In: *Esercizi Filosofici* 10.2 (2015), pp. 230–245.

— "Assertion among the speech acts". In: *The Oxford Handbook of Assertion*. Ed. by Sanford Goldberg. Oxford: Oxford University Press, 2020, pp. 159–178.

Schiffer, Stephen. *Remnants of Meaning*. Cambridge, Massachusetts: The MIT Press, 1987.

Schwenkler, John. *Anscombe's Intention: A Guide*. Oxford: Oxford University Press, 2019.

Shapiro, Lionel. "Commitment Accounts of Assertion". In: *The Oxford Handbook of Assertion*. Ed. by Sanford Goldberg. Oxford: Oxford University Press, 2020, pp. 75–98.

Sheridan, Richard Brinsley. *The Rivals*. 6th ed. London: G. Wilkie, 1798.

Smullyan, Raymond M. *Theory of Formal Systems*. Princeton, New Jersey: Princeton University Press, 1961.

Sperber, Dan and Deirdre Wilson. *Relevance: Communication and Cognition*. 2nd ed. Malden, MA: Blackwell Publishing, 1986/1995.

Stainton, Robert J. "A Deranged Argument Against Public Languages". In: *Inquiry* 59.1 (2016), pp. 6–32.

Stalnaker, Robert Culp. "Pragmatics". In: *Context and Content: Essays on Intentionality in Speech and Thought*. New York: Oxford University Press, 1970/1999, pp. 31–46.

— "Pragmatic Presuppositions". In: *Context and Content: Essays on Intentionality in Speech and Thought*. New York: Oxford University Press, 1974/1999, pp. 47–62.

— "Assertion". In: *Context and Content: Essays on Intentionality in Speech and Thought*. New York: Oxford University Press, 1978/1999, pp. 78–95.

Stalnaker, Robert Culp. "On the Representation of Context". In: *Context and Content: Essays on Intentionality in Speech and Thought*. New York: Oxford University Press, 1998/1999, pp. 96–113.

— *Context*. New York: Oxford University Press, 2014.

Stoecker, Ralf. "Action Explanation". In: *A Companion to Donald Davidson*. Ed. by Ernest Lepore and Kirk Ludwig. West Sussex: Wiley Blackwell, 2013, pp. 15–31.

Stojnić, Una. *Context and Coherence: The Logic and Grammar of Prominence*. Oxford: Oxford University Press, 2021.

Strawson, Peter Frederick. "Singular Terms, Ontology, and Identity". In: *Mind* 65.260 (1956), pp. 433–454.

Szabó, Zoltán Gendler. *Problems of Compositionality*. New York: Routledge, 2000.

Tomasello, Michael. *Origins of Human Communication*. Cambridge, Massachusetts: The MIT Press, 2008.

Townsend, David J. and Thomas G. Bever. *Sentence Comprehension: The Integration of Habits and Rules*. Cambridge, Massachusetts: The MIT Press, 2001.

Travis, Charles. "Sublunary Intuitionism". In: *Occasion-Sensitivity: Selected Essays*. New York: Oxford University Press, 1998/2008, pp. 130–149.

— "Truth and Merit". In: *The Philosophy of J. L. Austin*. Ed. by Martin Gustafsson and Richard Sørli. Oxford: Oxford University Press, 2011, pp. 175–203.

— "Views of my Fellows Thinking (Counting Thoughts)". In: *dialectica* 71.3 (2017), pp. 337–378.

Verhaegh, Sander. *Working from Within: The Nature and Development of Quine's Naturalism*. Oxford: Oxford University Press, 2018.

— "The Behaviorisms of Quine and Skinner: Genesis, Development, and Mutual Influence". In: *Journal of the History of Philosophy* 57.4 (2019), pp. 707–730.

Yalcin, Seth. "Semantics and Metasemantics in the Context of Generative Grammar". In: *Metasemantics: New Essays on the Foundations of Meaning*. Ed. by Alexis Burgess and Brett Sherman. New York: Oxford University Press, 2014, pp. 17–54.