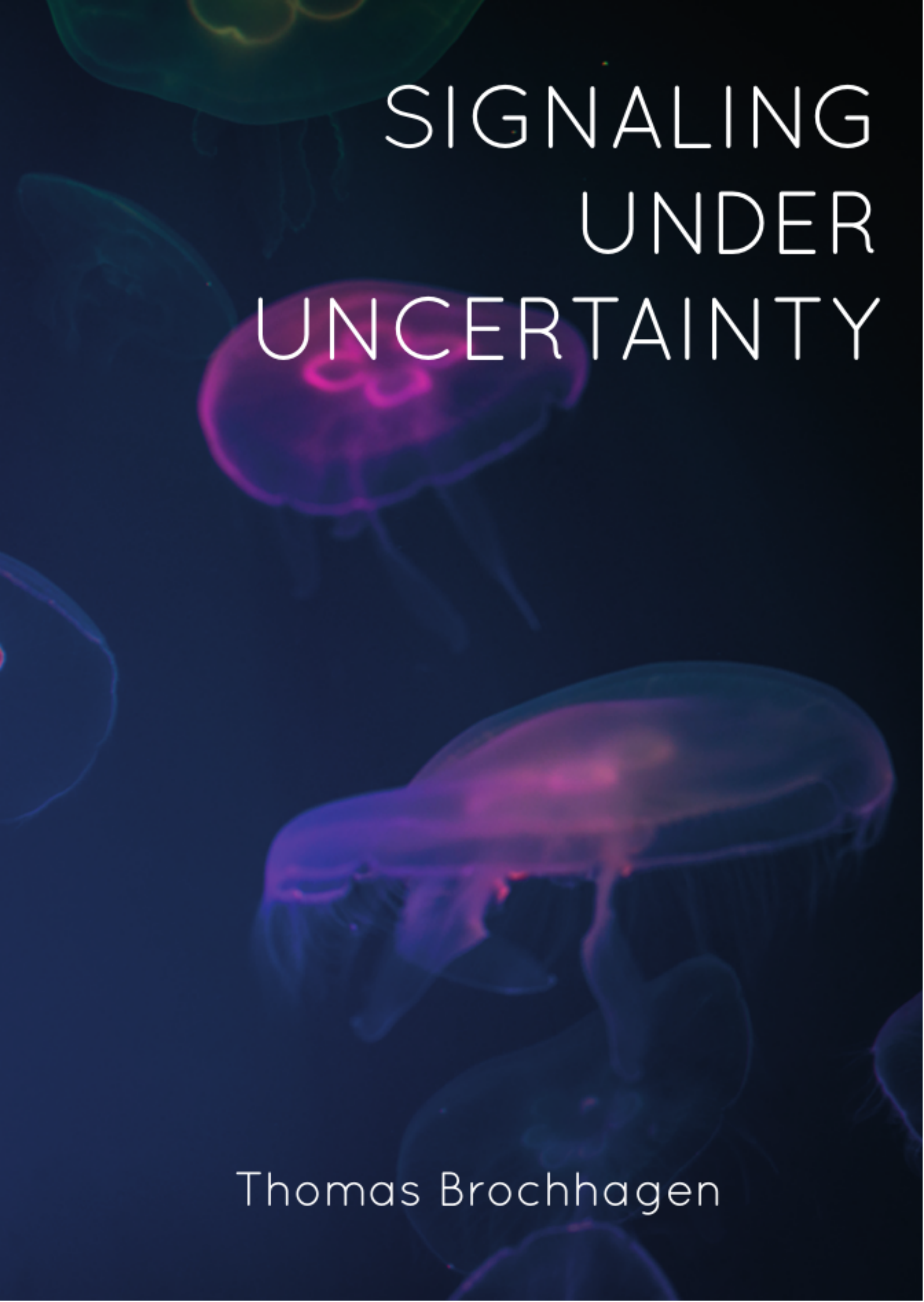


SIGNALING UNDER UNCERTAINTY

Thomas Brochhagen

Signaling Under Uncertainty

Thomas S. Brochhagen

Signaling Under Uncertainty

ILLC Dissertation Series DS-2018-06

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam
phone: +31-20-525 6051
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

This research was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 567652 /*ESSENCE: Evolution of Shared Semantics in Computational Environments*/

Cover design by Thomas S. Brochhagen
Printed and bound by Ipskamp Printing

ISBN: 978-94-028-0939-8

Signaling Under Uncertainty

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op woensdag 4 april 2018, te 10.00 uur

door

Thomas Simon Brochhagen

geboren te Santiago de Chile, Chili.

Promotiecommisie

Promotor:	Prof. dr. ing. R.A.M. van Rooij	Universiteit van Amsterdam
Promotor:	Prof. dr. E. Klein	The University of Edinburgh
Overige leden:	Prof. dr. M.J.B. Stokhof	Universiteit van Amsterdam
	Prof. dr. F.J.M.M. Veltman	Universiteit van Amsterdam
	Prof. dr. L.W.M. Bod	Universiteit van Amsterdam
	Dr. W.H. Zuidema	Universiteit van Amsterdam
	Prof. dr. G. Jäger	University of Tübingen
	Prof. dr. B. Skyrms	University of California, Irvine

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Contents

Acknowledgments

In his *The Analytical Language of John Wilkins* Jorge Luis Borges criticizes attempts to make universal classifications. He exemplifies the arbitrariness and conjectural nature of such classifications by conjuration of a Chinese encyclopedia that purports to give a universal animal taxonomy. In it, animals are divided into categories such as *belonging to the Emperor*, *fabulous*, *stray dogs*, *included in this categorization*, *et cetera*, or *those that look like flies when seen from afar*. Acknowledging, with the right words and in the right place, all those that have contributed, directly or otherwise, to shaping the content of this investigation, and me along the way, feels a little like writing entries in Borges' fictional encyclopedia myself. The good news is that I'm not, as Wilkins and others, trying to come up with a universal scheme. I will therefore take the liberty to let my subjective views seep through these initial pages, while accepting that a couple of paragraphs will not be able to capture all I would want to convey on these matters. I hope the reader will indulge me; it's not often that I write a book.

First, I wish to thank my supervisor Robert van Rooij. Throughout these years, Robert has given me the opportunity to prod and poke in many directions. Some of these efforts led to the material found in this thesis. As may be expected, others turned out to be less fruitful, or their outcomes were too remote to be included here. Irrespective of the result, I wish to thank Robert for the trust placed in me throughout this journey, as well as for guidance along it. Thanks for giving me time to explore and err, but also for (much needed) pragmatism when it came down to finishing projects.

Although, and particularly in light of the fact that, my research progressed in directions away from his and that we were neither at the same institution nor in the same country, I wish to thank Ewan Klein for his open ear and critical mind. I have good memories of coffee sessions in Edinburgh revolving around language identification in the limit, Popper and falsification, and noun-noun compounds (not necessarily all in one sitting). Thanks for sharing your constructive and critical views with me.

I also wish to extend my gratitude to the remaining, non-supervisory, members of my doctoral committee: Martin Stokhof, Frank Veltman, Rens Bod, Jelle Zuidema, Gerhard Jäger, and Brian Skyrms. I'm happy that you all accepted to be on board and really appreciate you devoting your time to my work.

I consider myself lucky to have met a number of people who have influenced me for the better in these past years. Among them, Liz Coppock stands out as the most formative person in my early academic life. Back when I was a master's student in Düsseldorf, she not only took me serious when I was but a fledgling but also actively encouraged me to do research with her. While I ended up working on very different topics in Amsterdam, Liz was the one who showed me the ropes, taught me how to write articles, and motivated me to move forward. Thank you for being a great mentor, colleague, and friend!

Another person that deserves a huge thank you in the category of big serendipitous influences (and nameless others) is Michael Franke. The impact Michael has had on my research is directly reflected in this thesis. Much of this material sprung either directly from our joint work or at least shows traces of discussions we had. Beyond your obvious academic footprint, I wholeheartedly thank you for inspiration, for your readiness to offer support, and for exemplary – always highly appreciated, I must add – meticulousness.

To the rows of students who have listened to what I had to say and questioned the questionable: thanks for your trust, kindness, and for keeping me on my toes. Fije van Overeem and Heidi Metzler, who embarked on longer individual research projects with me, deserve special mention: thanks for a great time and for the many lessons.

While it is often said that no man is an island, Amsterdam's relentless rain could at times suggest that I may become one; or that I may even drown in the process. I thank friends and colleagues from Amsterdam, Edinburgh, Trento, Paris, and Barcelona for keeping me afloat with discussion and distraction. Thanks, in particular, to great office mates and friends from the ILLC, who livened up the daily routine with sour coffee, food, football, bouldering, movies, and an occasional, I still believe misguided, optimism about the weather. A big shout-out to Paolo for debating choice under uncertainty with me when I needed it the most; to Arnold for his kind comments on parts of this thesis; to Nadine for the enduring distillery of my thoughts throughout these years and all the fun away from the office; and to Bastiaan for crafting and refining the Dutch summary of this thesis, as well as for much mischief. Gracias, gràcies e grazie to the Tordera household for providing me with a warm refuge to retreat to. I also would like to emphatically thank the ILLC office, in particular Jenny Batson, for years of support. I'm sure you have saved me from more than one headache.

Those across the ocean, from past and present, I thank for unrelenting support and care. I'm sorry for my intermittent radio silence and for having visited less than I would like. Thanks for always having my back. The same is true of my family. Un abrazo y gracias a la distancia a mis hermanos, Nicolás y Michael, a

Valentina, a mi madre Ana, und an meinen Vater Dieter.

Finally, it is impossible for a few words to do justice to what it means to me to have had Paula walk this journey with me. And not only did she walk it but she actively partook in it. Thank you for unwaivering support; for being there even when far away; for engaging with and commenting on my work throughout all its stages, from manuscripts to papers and from talks up to this entire thesis; for smiling with me; for making the writing of this paragraph very difficult in the most joyous way. A por muchas más sonrisas, perros, y revoluciones!

T.S. Brochhagen

Amsterdam, January 2018

Chapter 1

Introduction

> . . . In that empire, the art of cartography attained such perfection that the map of a single province occupied the entirety of a city, and the map of the empire, the entirety of a province. In time, those unconscionable maps no longer satisfied, and the cartographers guilds struck a map of the empire whose size was that of the empire, and which coincided point for point with it. The following generations, who were not so fond of the study of cartography as their forebears had been, saw that that vast map was useless, and not without some pitilessness was it, that they delivered it up to the inclemencies of sun and winters. In the deserts of the west, still today, there are tattered ruins of that map, inhabited by animals and beggars; in all the land there is no other relic of the disciplines of geography.
>
> Jorge Luis Borges, *On Exactitude in Science*

Communication is a social endeavor of information transfer. If we are told

(1) Alice went to Las Vegas and married,

we may learn just that. First, that Alice went to a place called Las Vegas. Second, that she married. However, we might also infer more. For instance, that Alice married in Las Vegas, taking *and* to indicate a temporal succession of events; or that *Las Vegas* refers to a famous place in Nevada rather than to the city of Las Vegas on the coast of Uruguay. With the appropriate background knowledge, we might even infer that Alice left her partner if the speaker is Alice's (now former) spouse.

Some of these inferences, such as that of Alice leaving her spouse, are rather ad hoc. Others, such as the enrichment of *and* to convey *and then*, show striking regularities across languages. What they all have in common is that they go beyond what is said explicitly. On the one hand, this can give rise to uncertainty and misunderstanding. Hearers cannot be certain that what they infer is intended, nor can speakers be certain that the inferences they intend to convey are drawn. On the other hand, the trait of not codifying all information overtly is not exclusive to natural language but found in much of biological signaling, from cellular communication to that of meercats and baboons (Greenough et al. 1998, Arnold and Zuberbühler 2006, Santana 2014). Rather than avoiding it, natural communication seems to thrive in the implicit; in the unsaid; in the contextually determined.

Framed in linguistic terminology, the information that is literally associated with an expression concerns its semantics. What is inferred beyond its literal meaning lies in the realm of pragmatics. This can involve the recruitment of contextual information, as well as mutual reasoning about interlocutors' linguistic choices. Under this distinction the meaning of an utterance is the product of conventional semantic meaning and general pragmatic rules that apply on language use in context.

Following the classic distinction between semantics and pragmatics we may then ask: if all interlocutors cared about was faithful information transfer, why leave to pragmatics and the implicit what semantics can do? Three reasons come to mind. First, it might be that misunderstandings are rare. What speakers intend to convey and what hearers take them to convey usually coincides. Second, it might be that some degree of uncertainty is unavoidable. After all, natural communication takes place in open and changing environments. Additionally, language is not acquired from a single source, nor does it serve a single purpose. It might consequently be impossible to use language in such a way that all uncertainty is quenched. Third, some degree of uncertainty might be advantageous. For instance, it may help interlocutors cope with some of the aforementioned issues, leaving to pragmatics the job of filling in gaps impossible to fill only by semantic conventions; or it might confer them with means to convey information in a more efficient manner. Inversely then, if one or a combination of these answers holds, we should also ask why and under which conditions interlocutors would leave to semantics what pragmatics can do.

The overall goal of this investigation is to address both of these questions by elucidating conditions under which language may come to leverage or accommodate uncertainty in information transfer. In particular, we will focus on cases in which speakers could, in principle, provide more information overtly but nevertheless often choose not to do so. In analogy to Borges' fictional empire, this investigation's underlying theme is accordingly the communicative potential that less (overt) exactitude offers in a trade-off against (pragmatic) uncertainty, as well as the linguistic properties that this trade-off gives rise to. Is language that leaves no room for uncertainty even a stable alternative, or would it be left in tatters by future language users?

1.1 The Semantics-Pragmatics Distinction

Natural languages are acquired from different sources and used in novel situations, often with new interlocutors of which little to nothing is known. As mentioned above, some variation across speakers and uncertainty about their language (use) may therefore be unavoidable. It is nevertheless also true that speakers do not necessarily shy away from, but regularly make use of expressions that invite or even necessitate pragmatic inference. A request for a blanket can be politely veiled by saying *I'm cold*; a temporal succession of events can be communicated by the order in which conjuncts appear, as in utterance (1); an invitation can be declined by saying *I have to work*. Crucially, such information could be conveyed more explicitly.

An influential account of the relation between what is said and what is conveyed is due to Grice (1975; 1989), who characterized pragmatic language use and its interpretation as resulting from a process of mutual reasoning about rational language use. That is, pragmatic inference is an outcome of a hearer's reasoning about why the speaker said what she said in the way she said it, taking into consideration the conversation's background as well as goals and beliefs of interlocutors. Conversely, a speaker reasons about her addressee's reasoning process, which she expects to effect a particular enrichment of her utterance. For instance, under the assumption that the speaker is cooperative and relevant, *I have to work* can be interpreted as providing a reason why the speaker will not be able to accept an invitation. Under this view, then, what is conveyed is explained in terms of the goals that language use is believed to serve. By contrast to many approaches in the philosophy of language contemporary to it, the Gricean project explicitly brings interlocutors, their goals, and the context of interaction into the picture instead of abstracting away from them.

Central to Grice's pragmatic theory is the notion of rationality. He embodies it in a number of guiding principles postulated to underlie conversation, his so-called *conversational maxims*. Roughly put, these principles state that rational speakers should be as informative but not more informative than necessary; that they should be truthful, relevant, and brief, but that they should avoid ambiguity. As an overarching principle, they should speak in such a way that the conversational goal is furthered. According to Grice, at a fundamental level this goal is to reach mutual understanding. These principles are not meant to be descriptive but normative (Grice 1989:§2:29). That is, they are not intended to describe how interlocutors behave but how rational language users ought to behave to reach mutual understanding. Pragmatic inferences then follow from the mutual assumption that all conversational participants behave in this fashion. What is more, not only the compliance with conversational maxims can give rise to pragmatic inference but also their violation. Under the assumption that (rational and cooperative) speakers try to comply with the maxims as much as they can, flouting a maxim is a deliberate and therefore meaningful signal for the hearer. In sum, rationality is seen as not only guiding, but also as constraining language use in relation to interlocutors' beliefs and goals (Westera 2017:6). Under this view, the role of semantics is to provide the groundworks on which pragmatic inference can build on.

Of course, although widespread, this is a particular view of pragmatics and its relation to semantics. Much research has been devoted to the explanatory potential of alternative principles to those proposed by Grice (e.g., Sperber and Wilson 1986, Wilson and Sperber 2006, Carston 2006), or their reduction and refinement (e.g., Horn 1984, Levinson 2000). As detailed in Chapter 2 we will follow a third way and ground notions such as cooperativity directly in the beliefs and preferences of interlocutors in context (e.g., Parikh 1991; 2000, Benz et al. 2006a, Benz 2006, Benz and van Rooij 2007, Franke 2009, Frank and Goodman 2012, Franke and Jäger 2014; 2016a). Under this view, pragmatic inference follows directly from reasoning about such contextual beliefs and preferences, without need for appeal to maxim-like rules. Using game-theoretic models that embody this view will enable us to inspect predictions borne out from an interactive perspective of language use, as well as those that follow from linguistic pressures that apply on such interactions.

The approach we take here is notwithstanding Gricean in spirit. Information transfer is viewed as en endeavor of social reasoning about rational language use. Schematically, we will view what is conveyed as a product of (cf. Parikh 2000):

$$\text{an agent's cognitive make-up} \otimes \text{context of utterance} \otimes \text{semantic meaning}$$

As a coarse approximation, our general explanandum can be recast as asking for the conditions that may favor information transfer that relies more strongly on the third component than on one of the first two, and vice-versa.

1.2 General Methodology

Our analysis spans across three interwoven levels: single interactions, iterated interactions, and the level of populations. As made precise in Section 2.4, linguistic behavior in single interactions is the foundation on which we build. Such behavior results from the context of interaction and an individual's cognitive make-up, her beliefs and preferences, the semantic conventions she holds to be true, and the conversational rules that she takes to operate on these conventions. Taken together, these factors determine agents' choice probabilities in production and comprehension in a given situation. However, the particular behavior of an agent at a particular time is not informative about the effects that linguistic pressures have on her language and behavior in the long run. Our central tenant is that if we are to understand why languages exhibit the properties they do, we should consider the tasks they fulfill over time, as well as pressures that apply on them. Many, if not arguably most, of these tasks are social endeavors that involve joint rather than independent action. Our focus will accordingly lay on iterated interactions and population-level dynamics. The former trace linguistic change over the course of a sequence of linguistic interactions. This kind of change can be conceived as taking place over the course of (possibly multiple) dialog(s). The latter trace change as a product of the expected outcome of repeated interactions of members of a population (horizontal change), as well as the effects of generational turnovers – when old population members are replaced by new ones (vertical change). The remainder of this chapter sets the stage for such an analysis by clarifying, in general terms, what we mean by words such as change, evolution and development; in which relation iterated and population-level dynamics stand; and on which level of analysis we operate.

1.2.1 Ontogeny and phylogeny; biological and linguistic change

The relation between horizontal and vertical language change bears similarity to the biological distinction between ontogeny and phylogeny. In broad strokes, ontogeny studies the development of an organism throughout its lifetime. Human ontogeny, for instance, spans from the ovum's fertilization across embryogenesis, infant and adolescent development, up to the development of the traits of fully matured adults. Phylogeny instead studies the evolution of species or populations throughout generations, tracing their development and relationship to one another.

The relationship between the development of an organism on the individual level and that of its phylum was regarded as a fundamental topic in evolutionary and developmental biology before the turn of the 20th century. A popular view on this matter is illustrated by Ernst Haeckel's famous theory of recapitulation, which holds that ontogeny recapitulates phylogeny (Haeckel 1866). In other words, Haeckel's hypothesis was that the individual development of an organism passes through stages that represent the development of its species, with ontogenetic stages representing the features of its adult ancestors.[1] The appeal of such a mechanistic view of an organism's ontogeny, viewed as a (con)sequence of its phyletic history, is evident in light of its historical context: theories of recapitulation attempted to gain insight into the past through the analysis of the present, with Mendelian genetics still to gain traction and to ultimately displace recapitulation. Nowadays a relationship between ontogeny and phylogeny under any strong interpretation of recapitulation is widely taken to be untenable. The influence of phylogeny on ontogeny as well as the role of other, at recapitulation's height unknown or disregarded, determinants turned out to be more complex than initially thought (see Gould 1977 for historical details).

What we learn from this snapshot of the history of biology is first and foremost that relating processes of individual development to macro-processes from which

[1] Whether individual development faithfully traverses all the stages of its phylum's history, merely resembles (some of) them, and to which degree this is supposed to apply to an organism as a whole or to its parts individually, allowing for temporal divergence in their development, were issues of active debate at the height of recapitulation's popularity. These details need not concern us here but see Gould 1977 for a historical overview.

they (partially) draw is often non-trivial. Caution is particularly called for in the face of seemingly intuitive parallels, as illustrated by the conclusions drawn from the ontogenetic expression of pharyngeal slits in human embryos to illustrate how humans pass through a developmental fish-like stage. Interestingly, precursors to recapitulation can be found in early theories of the origin of language (Danesi 1993). For instance, in the assumption that the language acquired by children deprived of linguistic input would correspond to a/the proto-language from which modern languages could have derived. As in the case of biology, a parallelism between linguistic change at an individual level and its historical development is appealing, for it would allow for a detailed inspection of its earlier stages in living specimens, so to speak. In the case of language evolution this issue is particularly pressing given that language "leaves no direct imprint in the fossil record" (Bolhuis et al. 2014:3). For the purpose of this investigation the origins of language itself are not of primary relevance. Our starting point is instead given by the change of pre-existing linguistic knowledge at different transmission levels with the goal to understand the conditions that lead to the adoption of linguistic strategies that may favor implicit over explicit information codification. Nevertheless, the question how the vertical transmission of linguistic knowledge affects its horizontal use and change, and vice-versa, is relevant here as well.

As with ontogeny and phylogeny, the emergence and change of language and its properties is also influenced by many intertwined factors. These range from biological and socio-ecological to cultural ones (Benz et al. 2006b, Steels 2011, Tamariz and Kirby 2016). Social and ecological pressures determine communicative needs, while biology determines the architecture that enables and constrains the means by which they can be fulfilled. Which of these factors is involved; whether change involves individual- or population-level processes; and on what timescale such change operates on are issues often obscured by the term *language evolution*. Let us therefore pause and briefly clarify these matters to set the scope of this investigation. With respect to the first issue concerning the nature of the described change, our focus will lay on cultural aspects. That is, we analyze processes of linguistic change as shaped by language use and its transmission: as a result of a process of cultural evolution (Christiansen and Chater 2008, Pagel 2009, Thompson et al. 2016). With respect to the second issue, drawing from the caution expressed above, we will analyze the effects of change at individual- and population-level separately, and contrast their outcome where pertinent. In analogy to the terminological distinctions often employed in connection to ontogeny and phylogeny, we refer to the former as (individual) *development* and reserve the term *evolution* for population change. Whether we analyze change at the individual- or population-level will depend on the phenomenon at hand. In Chapter 3, we will be concerned with contextual disambiguation in dialog. The inferences that resolve ambiguity in such cases can be rather ad hoc and idiosyncratic because they depend on the context and the interlocutors involved. Their treatment accordingly calls for models that make predictions about agents' choice in single interactions and track their change over repeated interactions. By contrast, Chapter 4 and Chapter 5 analyze the evolution of more systematic pragmatic inferences. This analysis abstracts from proximate causes, individual choices at particular points in time, and instead looks at the outcome of pressures that apply on populations of communicating agents. With respect to the third issue, as mentioned above, we constrain our attention to change effected by pressures such as ones for higher communicative success, learnability, or speaker-economy on populations or individuals that have some initial linguistic conventions to draw from; rather than their emergence, for example, from proto-communication systems, or the evolution of the cognitive endowment necessary to deploy pragmatic reasoning (see Woensdregt and Smith 2017 for a recent survey on these matters).

1.2.2 A computational analysis of outcomes of ecologically rational linguistic behavior

Marr (1982) famously argued for a tripartite distinction of analysis. His aim in doing so was to clarify how different perspectives taken toward an object of study are informed by different methodologies, and to clarify that they seek to answer different questions. More precisely, Marr proposes to categorize analysis according to the following complementary levels:

- **Computational level**: the what and why of a system/operation;
- **Algorithmic level**: the (computational) implementation of a system/operation. In particular, the representation of its input and output;
- **Implementational level**: the physical realization of a system/operation.

For example, in the case of vision Marr argues that a purely physiological description of its biological architecture may not necessarily add to our understanding of visual recognition. In particular, it may not add to our understanding of the motivations that underlie it; this being a computational rather than implementational question.

Of course, levels of analysis also interact and should therefore inform each other. Just as the physiology of vision may tell us something about its function, its computational description may guide its implementation. A transversal analysis is ultimately necessary to fully understand a complex system such as vision or, in our case, language; however impractical this task may be (Marr 1982:20).

Acknowledging at which level analysis is conducted has the advantage of constraining the perspective taken with respect to an object of study, as well as that of making clear the goals of the analysis. This is not only important to ensure internal coherence but also for critical assessment.

Our present aim is to gain insight in conditions under which language accommodates or leverages uncertainty "[...] by understanding the nature of the problem being solved [rather] than by examining the mechanism (and the hardware) in which it is embodied" (Marr 1982:27). Under Marr's classification, this investigation is then conducted at the computational level. We focus on two fundamental and interrelated problems being solved. The first is efficient information transfer through language use. The second is the transmission of linguistic knowledge from one agent to another. This may involve two proficient language users that adapt their language use to each other through the course of their interactions (Chapter 3), or proficient language users from which naïve users learn (Chapter 4 and 5). Put differently, the second problem concerns the acquisition or adaptation of the means by which the first problem is solved. As we shall see, solutions to these problems can pull in opposite directions. A characterization of their joint influence and combined solution is therefore part of our overall goal.

With Grice and much work in Bayesian cognitive modeling, decision theory, and game theory, our approach is rationalistic at the level of individuals (Anderson 1990; see Griffiths et al. 2012 and Franke and Jäger 2016a for discussion). This means that we aim to give a teleological, rather than mechanistic, explanation of linguistic behavior. To analyze linguistic change, we couch this rationalistic approach in the ecological context in which behavior takes place. That is, we analyze linguistic change as shaped both by the behavior resulting from the computational capacities of an agent itself, as well as by the environment in which this behavior is embedded (Simon 1990). The former we assume to correspond to (an approximation of) bounded rational behavior (Chapter 2). The latter encompasses factors such as the interlocutor's overt behavior and contextual information (Chapter 3 and 5), the population in which actors find themselves in (Chapter 4 and 5), as well as factors such as noisy perception (Chapter 6). In light of our main findings, we cast a critical light on this approach to the analysis of linguistic change in Chapter 6.

1.3 Overview and Source Material

Chapter 2 This chapter lays out the technical and conceptual foundations of our analysis, building on Lewis' (1969) signaling games. We proceed by incrementally introducing some central game-theoretic notions and highlight how they can aid linguistic inquiry. In particular, we focus on how they can make the interplay of conventional meaning, interlocutors' goals, information transfer, and mutual reasoning precise.

This chapter also discusses the limitations of static equilibrium analysis. With Franke (2013) and Huttegger and Zollman (2013), we argue that static approaches suffer from conceptual and technical issues that make them unsuitable for our purposes: they fail to make clear predictions when multiple equilibria exist; their procedural agnosticism lacks in explanatory force to address the question how language users may come to adopt a particular (way of using) language; and they can be taken to suggest outcomes that in some cases are seldom, if ever, reached. These shortcomings motivate a move from a static analysis of language to a dynamic one. This is the kind of analysis which we conduct throughout this investigation.

In the dynamic realm we differentiate between micro dynamics, which track change in language or behavior of individual agents, and macro dynamics, which abstract away from individuals and instead trace change in populations. Making predictions using either type of dynamic analysis presupposes that we characterize how language is used, as well as what counts as a language in the first place. To this end, we introduce a general model of rational language use at this chapter's end.

Chapter 3 This chapter focuses on ambiguity in iterated interactions. In particular, on the question why ambiguity is such a pervasive property in biological signaling if, at first sight, functional considerations about efficient and accurate information transfer would seem to disfavor it. With previous justifications of ambiguity, we argue that context plays an important role in allowing for the (relatively) safe exploitation of ambiguity. However, we inject some wrinkles in this justification by calling into question the assumption that interlocutors have access to the same contextual information to disambiguate utterances. We then argue that this issue unravels into a larger one, where the interaction between context, interlocutors' private contextual expectations, and their beliefs about each other's expectations play an important role. These factors are argued to jointly determine the conditions under which a functional advantage for ambiguity crystallizes. We conclude that ambiguity can be viewed as an opportunistic adaptive device: it endows interlocutors with the ability to flexibly mold language use to suit their communicative preferences and the context of interaction.

Iterated interactions and alignment play an important role in this chapter. By interacting multiple times, interlocutors can learn something about each other's contextual expectations. This reduces the speaker's uncertainty about what her interlocutor is likely to infer from an ambiguous utterance.

We analyze the outcomes of iterated interactions without a common contextual prior using a conservative generalization of previous models of rational language use, paired with simple update rules. After exploring the theoretical predictions of the model, we show that it succeeds in explaining signaling patterns found in experimental data.

The content of this chapter is based on:

Brochhagen, Thomas. 2017. Signalling under uncertainty: interpretative alignment without a common prior. *The British Journal for the Philosophy of Science.* doi: 10.1093/bjps/axx058.

Chapter 4 This chapter focuses on the evolution of a division of labor between semantics and pragmatics. To analyze how such a division may come to be, we trace the effects that two evolutionary pressures have on the joint interaction between conventionalized lexical representations and conversational strategies of language use. These pressures are (i) a horizontal one for communicative success during information transfer within a population and (ii) a vertical one for learnability, which applies when linguistic knowledge is transmitted from one generation to the next. We model the ensuing dynamics using the replicator-mutator dynamic, where replication exerts fitness-based pressure for efficient communication and mutation captures the transmission fidelity by which linguistic knowledge is transmitted through a process of iterated learning. Importantly, learners do not have access to unobservable lexical representations and conversational strategies. They instead need to infer these latent properties from the overt linguistic behavior that results from their combination.

We analyze the separate and joint influence that these pressures have in a case study on the (lack of) lexicalization of scalar implicatures. This case study suggests that semantics and pragmatics play a synergic role in overcoming both pressures: pragmatic use allows maintenance of simpler lexical representations that are easier to learn; pressure toward representational simplicity indirectly promotes pragmatic over literal language use. As a consequence, iterated transmission and use of language lead to a regularization that may explain the lack of lexicalization of systematic pragmatic enrichments.

This chapter is based on:

Brochhagen, Thomas, Michael Franke and Robert van Rooij. Co-evolution of lexical meaning and pragmatic use. 2017. *Manuscript*, Amsterdam–Tübingen.

Brochhagen, Thomas, Michael Franke and Robert van Rooij. 2016. Learning biases may prevent lexicalization of pragmatic inferences: a case study combining iterated (Bayesian) learning and functional selection. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

Chapter 5 This chapter looks at ambiguity at the population level. Drawing from the individual-level analysis in Chapter 3, we ask under which conditions conventional semantic meaning that allows for functional ambiguity exploitation evolves. For signaling behavior to be functionally advantageous it needs to ensure that information is transmitted accurately. This means that, even if a signal is semantically ambiguous, in context it should be, by and large, unequivocal. However, such a signal may not necessarily suggest underlying semantic ambiguity to a naïve learner. If the learner only witnesses the signal being used in a single context to signal a single meaning, then she may not learn to associate this signal with other meanings. This poses a challenge for the acquisition of (unobservable) ambiguous semantic conventions.

We use the model from Chapter 4 to investigate how the context(s) in which communication and learning take place affect the evolution of semantic ambiguity. Our results suggest that ambiguity evolves when the environment is varied, with language use happening in multiple contexts that are informative about different meanings. An environment that instead favors a single context promotes precise semantic conventions rather than the pragmatic flexibility enabled by their underspecification.

Chapter 6 This chapter discusses the models proposed in previous chapters and the predictions they make from a general perspective. We begin by reflecting on what we learned about the conditions under which language may come to favor semantic underspecification and recruit pragmatics to effect efficient and successful information transfer. We argue that there are multiple evolutionary trajectories under which this may happen. First, if communication occurs in varied informative contexts, then underspecified semantics coupled with pragmatic abilities endow interlocutors with the ability to flexibly adapt their linguistic resources to the context of interaction and their interlocutors. Second, some underspecified lexical meanings may be simpler and therefore easier to learn; if interlocutors are sufficiently rational, then pragmatic reasoning can enrich these meanings and thereby counteract functional disadvantages otherwise incurred. Reversely, if the context of interaction is static or rationality is low, then precise semantics come to be favored over pragmatic recruitment. We then discuss the methodological issues raised by this kind of investigation and argue for a pluralistic approach that takes multiple likely factors of change into consideration.

This chapter discusses results presented in fuller detail in:

Brochhagen, Thomas and Michael Franke. 2017. Effects of transmission perturbation in the cultural evolution of language. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

Chapter 7 This is where we conclude. This chapter gives a broad summary of our findings and a sketch of roads ahead.

Chapter 2

Signaling Games: Analysis and Interpretation

> "I don't know what you mean by 'glory'," Alice said. Humpty Dumpty smiled contemptuously. "Of course you don't – till I tell you. I meant 'there's a nice knock-down argument for you!' "
> "But 'glory' doesn't mean 'a nice knock-down argument'," Alice objected. "When I use a word," Humpty Dumpty said, in rather a scornful tone, "it means just what I choose it to mean – neither more nor less."
>
> Lewis Carroll, *Through the Looking-Glass*

Where our goal is to analyze the conditions that may give rise to linguistic properties in interaction, we first need to specify how the choices that make up such interactions are made. Specifying linguistic choice, in turn, requires the specification of interlocutors' communicative goals, the context of interaction, and other aspects that may influence how these goals are reached. For instance, interlocutors' beliefs about each other's linguistic behavior. Game theory gives us the means to make these notions and their interplay precise.

The fundamentals of game theory were laid out in von Neumann and Morgenstern's (1944) *Theory of Games and Economic Behavior.* In it, a game is understood as any interaction between agents for which all possible actions and their joint outcome can be specified. A straightforward case that satisfies these conditions is an interaction with simple and overtly acknowledged rules, actions, and goals, such as a game of rock-paper-scissors. Table 2.1 shows this game in normal form, which is read as follows. The possible moves of one player, call her player one, are represented by the table's rows. The possible moves of the other player, call her player two, are given by the table's columns. According to this specification each player can perform one of three actions: *rock*, *paper* or *scissors.* The rules of the game dictate that both players should reveal their choices simultaneously. The outcome of the combination of their choices is described by the

		Player 2		
		rock	paper	scissors
Player 1	rock	$(0,0)$	$(-1,1)$	$(1,-1)$
	paper	$(1,-1)$	$(0,0)$	$(-1,1)$
	scissors	$(-1,1)$	$(1,-1)$	$(0,0)$

Table 2.1: Normal-form representation of a game of rock-paper-scissors.

table's cells, where a numeric value of 1 is attached to winning, -1 to losing, and 0 to a draw. In Table 2.1 the payoff of player one is given by the first number in a cell's bracket and that of player two is given by the second number. In this way, we know that the pair of actions ⟨*rock,rock*⟩ gives a payoff of $(0, 0)$, a draw, whereas ⟨*rock,scissors*⟩ is a win for player one and a loss for player two.

Even for this seemingly mundane game there are some interesting questions a game theorist may want to answer. For instance, one may ask what action a player should take given the information she has about her opponent's behavior. If it is known, for example, that the opponent plays *scissors* half of the time. Another question one may ask is how an action policy should change over time. For example, after witnessing the opponent play *rock* 10 times in a row.

The broad conception of a game put forth by von Neumann and Morgenstern allows us to ask similar questions about other kinds of interactions. Indeed, game theory has found applications in a multitude of fields, ranging from economics, political science and biology to computer science and linguistics. In this chapter we discuss some of the questions that the conception of language (use) as a game can inform us about, as well as how we may go about answering them.

Section 2.1 doubles as an introduction to fundamental game-theoretic notions as well as to Lewis' (1969) signaling games. Section 2.2 builds on these notions to characterize linguistic outcomes, using classic static solution concepts such as that of a Nash equilibrium. The shortcomings of static equilibrium analysis are discussed in detail in Section 2.3, and contrasted with dynamic analysis. This discussion motivates the kind of analysis we conduct throughout this investigation. Section 2.4 introduces a family of models of rational language use, which we employ in following chapters to characterize linguistic behavior. In particular, to characterize pragmatic inference. Final remarks on our main assumptions are given in Section 2.5.

2.1 Signaling Games

In his seminal work on conventions, Lewis (1969) laid out the backbone on which much of modern game-theoretic analysis of communication rests (see Skyrms 2010 for an overview). According to Lewis communication can conceived as a strategic endeavor between interlocutors. Central to his analysis are *signaling games*: formal characterizations of information transfer mediated by messages sent from speakers to addressees. The classical setup considers two agents: a sender and a receiver. The sender has some information that she wishes to convey to the receiver. As the receiver has no direct access to the sender's information state, the sender has to resort to the use of messages. Upon reception of a message the receiver's task is to act upon it.

Signaling games can characterize a variety of communicative situations. For instance, animal alarm calls. In the signaling literature, vervet monkeys are particularly famous for their alarm calls (e.g., Seyfarth et al. 1980, Cheney and Seyfarth 1990, Skyrms 2010:§2, Price et al. 2015; see Zuberbühler 2009 for an overview on animal alarm calls). A vervet alarm call depends on the type of predator observed. This allows receivers of the alarm call to take appropriate evasive action. For example, the alarm call for an aerial predator may effect the act of hiding in bushes. One used for terrestrial predators may instead be answered by hiding in trees. By analogy, in the case of human communication, the appropriate act to a request such as *Pass me the salt* would be to pass the salt (if the addressee is able and willing to do so). The particular kind of receiver acts we will focus on in this investigation are interpretations. That is, we assume the receiver's act to be to interpret the message she receives as a particular state. This simplifies our notation a little, as we need not focus on sender states and receiver acts, but rather on a single set of states relevant to the context of interaction. More precisely, a classical signaling game considers a set of sender states S, a set of messages M, and a set of receiver acts A. An interpretation game is one where $A = S$.

The strategic aspect of interactions in signaling games lies in the choices made by each agent and the joint outcome they wish to effect. What message the sender sends for which state hinges on the receiver's (expected) interpretation of the message. Conversely, the receiver's interpretation hinges on the way in which messages are (expected to be) used by the sender. If interlocutors can coordinate in such a way that messages sent in a state are interpreted as conveying that state then information is transferred faithfully.

Strategies. More formally, a signaling game is a sequential two player game. In contrast to a game of rock-paper-scissors, this means that choices are not simultaneous. Instead, sender choices are contingent on states (they are in) and receiver choices are contingent on messages (they receive). In other words, a receiver's interpretation follows a sender's choice. A player's complete contingency plan of which message/state to send/infer when is called a strategy. If these choices are deterministic, a sender strategy is a mapping from states to messages, $\sigma\colon S \to M$, and a receiver strategy is one from messages to states, $\rho\colon M \to S$. Such deterministic strategies are called pure in game-theoretic parlance.

Sender strategies				Receiver strategies			
σ_1:	$s_1 \mapsto m_1$ $s_2 \mapsto m_1$	σ_2 :	$s_1 \mapsto m_1$ $s_2 \mapsto m_2$	ρ_1:	$m_1 \mapsto s_1$ $m_2 \mapsto s_1$	ρ_2 :	$m_1 \mapsto s_1$ $m_2 \mapsto s_2$
σ_3:	$s_1 \mapsto m_2$ $s_2 \mapsto m_2$	σ_4 :	$s_1 \mapsto m_2$ $s_2 \mapsto m_1$	ρ_3:	$m_1 \mapsto s_2$ $m_2 \mapsto s_2$	ρ_4 :	$m_1 \rightarrow s_2$ $m_2 \mapsto s_1$

Table 2.2: Pure strategies in a 2-states/messages signaling game.

$$\begin{tikzcd}
s_1 \arrow[r] & m_1 \arrow[r] & s_1 \\
s_2 \arrow[ur] & m_2 \arrow[ur] & s_2
\end{tikzcd}
\qquad
\begin{tikzcd}
s_1 \arrow[r] & m_1 \arrow[r] & s_1 \\
s_2 \arrow[r] & m_2 \arrow[r] & s_2
\end{tikzcd}$$

Figure 2.1: Sequential depiction of σ_1 and ρ_1 (left) and σ_2 and ρ_2 (right).

We will also want to consider strategies that are not pure but instead capture probabilistic behavior. A probabilistic sender strategy is a mapping from states to probability distributions over messages, $\sigma\colon S \rightarrow \Delta(M)$. A probabilistic receiver strategy is one from messages to distributions over states, $\rho\colon M \rightarrow \Delta(S)$. We will often denote the probability of a sender choosing message m given state s under behavioral strategy σ as $\sigma(m \mid s)$. Analogously, $\rho(s \mid m)$ is the probability of the receiver interpreting message m as state s under ρ.

At first sight, it may seem strange for interlocutors to adopt a probabilistic strategy over a pure one. After all, while there are situations in which players gain from unpredictability, such as when playing rock-paper-scissors, less predictability in communication might work against mutual understanding. Mutual understanding improves if the receiver can reliably infer what the sender wishes to convey and the better the sender can predict the receiver's interpretative behavior. There are two general reasons why considering probabilistic strategies is nevertheless desirable. First, communication often involves a substantial amount of uncertainty, meaning that neither sender nor receiver can be certain about each other's behavior. Probabilistic strategies can accordingly be conceptualized as resulting from conjectures about interlocutor behavior (Franke 2009:§1.2.3). Second, as made precise in Section 2.4, probabilities are convenient to represent an agent's occasional deviation from rational behavior, failure to recognize slight differences between the outcome of different choices, other consequences of bounded rationality, or imperfect perception of the environment (see Chapter 6).

To make matters more concrete, consider a signaling game with two states, s_1 and s_2, and two messages, m_1 and m_2. In this game there are four pure sender strategies and four pure receiver strategies, listed in Table 2.2. The sequential nature of the game and the interdependence of sender and receiver strategies is illustrated in Figure 2.1.

Preferences and (expected) utility. Intuitively, some of the strategies listed in Table 2.2 are less suited for communication than others. For instance, a sender adopting strategy σ_1 will always send message m_1, irrespective of the state she is in. Conversely, a receiver using ρ_1 will interpret any message she receives as state s_1. By contrast, the strategy combinations $\langle \sigma_2, \rho_2 \rangle$ and $\langle \sigma_4, \rho_4 \rangle$ ensure that the state that the sender is in is also the one inferred by the receiver.

That sender and receiver care about matters such as successful information transfer is captured by the notion of *utility*, a subjective measure of an agent's preference over outcomes of a game. More precisely, utility is a function from outcomes to real numbers: $U_{\sigma,\rho} : S \times M \times S \to \mathbb{R}$.

Beyond having a preference about which state is inferred when, interlocutors can also have preferences about *how* matters are communicated. A sender may prefer a polite but less clear expression over one that is more explicit; have particular stylistic predispositions; prefer shorter over longer expressions; or have preferences over matters such as the relative cognitive load of the retrieval of expressions. A receiver may have other, possibly opposed, preferences over messages.

Following the distinction between what is communicated and how it is communicated, signaling games are often distinguished depending on whether players have differential preferences over messages. Signaling games in which one message is as good as any other to all players are known as *cheap talk*, so called because no message carries cost (or all are equally costly). By contrast, if preferences over messages are relevant, message cost is represented as a small but non-negligible amount deducted from an interlocutor's preference for the communicated state when the message is used. In other words, message cost is inverse to message preference but nominal. That is, small relative to preferences over what is communicated when (Blume et al. 1993, Benz and van Rooij 2007).

Even if signaling is free of cost, communicative preferences may be opposed. A sender could, for instance, prefer the receiver to always infer state s_1, irrespective of whether this is the actual state. The receiver might instead prefer to know the actual sender state. This situation could represent the preferences of an applicant who wants her interviewer to think she is qualified for a position even when she is not; that of a predator mimicking a harmless species to lure in prey; or that of prey mimicking a noxious species to avoid predation. In cases where players' preferences are orthogonal to each other communication is not a cooperative affair. If, as standardly assumed, players' preferences are common knowledge, a player's best interest would then be to detect deceitful behavior and turn it to their advantage. Whether messages are credible and information transfer is possible at all will then depend on how aligned preferences are, what each message is (believed) to mean, what other messages are at an agent's disposition, as well as other aspects of the interaction, such as message cost. In this investigation, we will instead be concerned with *cooperative* communication, meaning that sender and receiver strive for mutual understanding. In signaling games this is reflected by

		Receiver	
		s_1	s_2
Sender	s_1	1	0
	s_2	0	1

Table 2.3: Players' preferences in a cheap talk 2-states/messages signaling game.

utility functions that assign higher utility to successful information transfer than to misunderstanding, while leaving room for potentially diverging preferences over messages.

Taking stock, outcomes in signaling games are triples of a sender state s, a sent message m, and receiver interpretation s'. Letting $c_{\sigma,\rho}(\cdot)$ codify the subjective cost assigned to a message by sender σ or receiver ρ, their utility can be defined as:

$$U_\sigma(s,m,s') = \delta(s,s') - c_\sigma(m); \tag{2.1}$$

$$U_\rho(s,m,s') = \delta(s,s') - c_\rho(m), \tag{2.2}$$

where

$$\delta(s,s') = \begin{cases} 1 & \text{if } s = s' \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

In Gricean terms, $\delta(\cdot)$, as defined in (2.3), codifies cooperativity. In line with Grice's (1975) postulated maxims of rational language use, it moreover follows from rational choice as utility maximization that a sender will convey matters, e.g., as clearly but succinctly as possible while avoiding false or misleading statements (as long as individuals actually have such preferences). This game-theoretic rendering can accordingly capture the fundamental insights that underlie the Gricean program, but also allows us to go beyond it by enabling for the consideration of non-cooperative situations, as well as more differentiated preferences. We return to this issue and pragmatic inference more generally in Section 2.4.

If talk is cheap, preferences over outcomes in a cooperative interpretation game reduce to pairings of a sender state s and a receiver's interpretation s'. In a cooperative game these preferences are equal for both players. For a signaling game with two states they can be summarized by the matrix in Table 2.3.

Utility captures the quantitative preference of an agent for a single outcome. Expected utility takes this a step further and gives the weighted mean utility that interlocutors can expect when interacting. For finite S and two players, σ and ρ,

	ρ_1	ρ_2	ρ_3	ρ_4
σ_1	.5	.5	.5	.5
σ_2	.5	**1**	.5	0
σ_3	.5	.5	.5	.5
σ_4	.5	0	.5	**1**

Table 2.4: $\text{EU}(\sigma_i, \rho_j)$ in a cheap talk 2-states/messages signaling game.

expected utility is defined as:[1]

$$\text{EU}_\sigma(\sigma,\rho)=\sum_s P^*(s)\ \sum_m \sigma(m|s)\ \sum_{s'} \rho(s'|m)\ U_\sigma(s,m,s'); \tag{2.4}$$

$$\text{EU}_\rho(\sigma,\rho)=\sum_s P^*(s)\ \sum_m \sigma(m|s)\ \sum_{s'} \rho(s'|m)\ U_\rho(s,m,s'), \tag{2.5}$$

where $P^*(s)$ is the probability of state s. How P^* is to be interpreted in linguistic terms has been subject to some discussion (Allott 2006, Franke 2013) and will become relevant in later chapters. To better explain what is at stake we defer this issue to Section 3.2.2. We may until then think of P^* as an abstract exogenous determinant for what state the speaker is in, often referred to as *nature*.

2.2 Static Concepts and Optimal Solutions

With these notions at hand we can return to our 2-states/messages signaling game and quantitatively compare strategy pairings. For expository ease, let us assume that talk is cheap, that each state is equally probable, $P^*(s_1) = 1/2 = P^*(s_2)$, and focus only on pure strategies. The expected utilities of this game's 16 possible pure strategy combinations are given in Table 2.4.

Inspecting how well any two strategy pairings of this game fare reveals two things. First, many pairings leave no room for improvement. For instance, should a receiver be confronted with a sender that signals according to σ_1 then she can do no better than 0.5 under any strategy. To see this, inspect the row of σ_1 in Table 2.4. Put differently, a receiver following ρ_3, for example, has no incentive to change her interpretative behavior when interacting with σ_1. There is no alternative strategy that would yield better results. Second, the strategy combinations $\langle\sigma_2,\rho_2\rangle$ and $\langle\sigma_4,\rho_4\rangle$ guarantee the highest expected utility of 1.

[1] For pure strategies expected utility is defined by the more succinct:

$$\text{EU}_\sigma(\sigma,\rho)=\sum_s P^*(s)\ U_\sigma(s,\sigma(s),\rho(\sigma(s)));$$

$$\text{EU}_\rho(\sigma,\rho)=\sum_s P^*(s)\ U_\rho(s,\sigma(s),\rho(\sigma(s))).$$

$$\begin{tikzcd}
s_1 \arrow[r] & m_1 \arrow[r] & s_1 & & s_1 \arrow[dr] & m_1 \arrow[dr] & s_1 \\
s_2 \arrow[r] & m_2 \arrow[r] & s_2 & & s_2 \arrow[ur] & m_2 \arrow[ur] & s_2
\end{tikzcd}$$

Figure 2.2: Signaling systems of a 2-states/messages signaling game.

That is to say that either pairing guarantees that states are always communicated successfully. In contrast to the first case, there is not only nothing to be gained from unilaterally adopting a different strategy but doing so would always be detrimental.

Both of these properties are central in game theory. The former, when there is no incentive to unilaterally switch to a different strategy, is called a (*weak*) *Nash equilibrium*. The latter, when no incentive for unilaterally switching exists and, moreover, doing so would always be disadvantageous, is called a *strict Nash equilibrium*.

More formally, let s_i denote player i's strategy and s_{-i} the strategies of all players except i. Then $\left\langle s_i^*, s_{-i}^* \right\rangle$ is a weak or, respectively, strict Nash equilibrium should there be no alternative strategy s_i for any player i such that

$$\mathrm{EU}_i(s_i^*, s_{-i}^*) \leq \mathrm{EU}_i(s_i, s_{-i}^*); \tag{2.6}$$

or

$$\mathrm{EU}_i(s_i^*, s_{-i}^*) < \mathrm{EU}_i(s_i, s_{-i}^*). \tag{2.7}$$

To reiterate in words, a Nash equilibrium is a collection of strategies in which no single player has an incentive to change her strategy provided everyone else conforms to theirs. In the context of signaling games, Lewis (1969) calls those Nash equilibria that are strict and lead players to associate each single state with a single and correct interpretation *signaling systems*.

Central to a signaling system is that it is arbitrary. For the games Lewis focused on – cheap talk games with $|M| = |S|$ $(= |A|)$ and uniform P^* – there always exists at least one other equilibrium that results from a permutation of messages that is as optimal for information transfer as itself. Such is the relation between $\langle \sigma_2, \rho_2 \rangle$ and $\langle \sigma_4, \rho_4 \rangle$, depicted in Figure 2.2. Signaling systems consequently do not require messages to be meaningful in themselves. After all, messages can be used to signal completely different states and either signaling system is equally good for information transfer. Instead, what endows messages with meaning is their use in equilibrium because interlocutors behave "as if" they meant something. Moreover, although a signaling system is arbitrary, being a strict Nash equilibrium ensures that no individual would wish to unilaterally deviate from it.

As shown in Table 2.4, there are 2 signaling systems in a 2-states/messages game. More generally, in a N-states/messages signaling game, there are $N!$ signaling systems.

Signaling systems intuitively correspond to desirable linguistic outcomes. However, the characterization of strategy combinations in terms of equilibria is silent on how they are reached. This begs the question whether they can be reached at all; and if so, under which conditions. One common conception of Nash equilibria is that they correspond to stable states when games are repeatedly played (Osborne 2004). Under this view, equilibrium analysis is agnostic about the process that leads players to adopt their behavior but predicts that, over time, players will reach equilibrium and thereafter remain in it. Players in equilibrium stay in equilibrium. Importantly, equilibrium analysis does not appeal to the rationality of players in reaching a particular outcome.

Populations and types. Nash equilibria describe optimal stable outcomes with respect to individual strategy pairings. To transfer this idea to a population of communicating agents we either have to restrict our attention to senders only meeting receivers and vice-versa, i.e., consider two distinct populations and their interaction, or alternatively endow agents with both a sender strategy and a receiver strategy. The latter is an arguably natural assumption for much of biological communication and is what we will assume throughout this investigation.

Irrespective of the nature of the population analysis employed, we call the units that distinguish members of a population *types*, $\tau \in T$. Under a biological interpretation types can be identified with phenotypes and their expected utility with their fitness. This determines a type's chance of survival and reproduction. Under a cultural interpretation a type corresponds to a particular linguistic behavior. That is, a sender and receiver strategy. As exemplified by monkey alarm calls (e.g., Seyfarth et al. 1980, Cheney and Seyfarth 1990, Skyrms 2010:§2), communicative success can be a determinant for survival and reproduction, closing the gap between communicative and biological fitness; being able to act upon an alarm call correctly can greatly enhance chances of survival. However, where the focus lies on human communication, my preferred interpretation is that agents themselves strive toward efficient and successful information transfer. They therefore occasionally adapt or revise their behavior to improve their communication with other members of the population (Benz et al. 2006b:§3.3, Skyrms 2010:55). In a dynamic setting a type's higher fitness then translates to a higher chance that other agents will attempt to adopt/imitate this behavior (see Chapter 4 for details).[2]

The transformation of a game to one in which all players draw from the same strategy pool is called its symmetrization (Wärneryd 1993, Cressman 2003,

[2] Another possible cultural interpretation, put forth by Nowak et al. (2002), is that successful communication increases the chances of influencing the acquisition process of future generations. In the same fashion as my preferred interpretation, populations would come to reflect a higher proportion of successful past behaviors (assuming all types are equally likely to be acquired; see Chapter 4). Rather than this behavior being adopted by more agents horizontally, this would then be a consequence of the influence of their success on future generations.

	τ_{11}	τ_{12}	τ_{13}	τ_{14}	τ_{21}	τ_{22}	τ_{23}	τ_{24}	τ_{31}	τ_{32}	τ_{33}	τ_{34}	τ_{41}	τ_{42}	τ_{43}	τ_{44}
τ_{11}	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
τ_{12}	.5	.5	.5	.5	.75	.75	.75	.75	.5	.5	.5	.5	.25	.25	.25	.25
τ_{13}	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
τ_{14}	.5	.5	.5	.5	.25	.25	.25	.25	.5	.5	.5	.5	.75	.75	.75	.75
τ_{21}	.5	.75	.5	.25	.5	.75	.5	.25	.5	.75	.5	.25	.5	.75	.5	.25
τ_{22}	.5	.75	.5	.25	.75	**1**	.75	.5	.5	.75	.5	.25	.25	.5	.25	0
τ_{23}	.5	.75	.5	.25	.5	.75	.5	.25	.5	.75	.5	.25	.5	.75	.5	.25
τ_{24}	.5	.75	.5	.25	.25	.5	.25	0	.5	.75	.5	.25	.75	**1**	.75	.5
τ_{31}	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
τ_{32}	.5	.5	.5	.5	.75	.75	.75	.75	.5	.5	.5	.5	.25	.25	.25	.25
τ_{33}	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
τ_{34}	.5	.5	.5	.5	.25	.25	.25	.25	.5	.5	.5	.5	.75	.75	.75	.75
τ_{41}	.5	.25	.5	.75	.5	.25	.5	.75	.5	.25	.5	.75	.5	.25	.5	.75
τ_{42}	.5	.25	.5	.75	.75	.5	.75	**1**	.5	.25	.5	.75	.25	0	.25	.5
τ_{43}	.5	.25	.5	.75	.5	.25	.5	.75	.5	.25	.5	.75	.5	.25	.5	.75
τ_{44}	.5	.25	.5	.75	.25	0	.25	.5	.5	.25	.5	.75	.75	.5	.75	**1**

Table 2.5: $\text{EU}(\tau_i, \tau_j)$ in a symmetrized cheap talk 2-states/messages game.

Franke and Wagner 2014). While our illustrative 2-states/messages signaling game has 4 pure sender strategies and 4 pure receiver strategies, there are 16 types in its symmetrized counterpart. These result from pairing all possible sender and receiver strategies. Mnemonically labeling each type with an index of its sender and receiver strategy we have that $\tau_{11} = \langle \sigma_1, \rho_1 \rangle$ and that $\tau_{42} = \langle \sigma_4, \rho_2 \rangle$, for example. That is, τ_{11} sends and receives according to σ_1 and ρ_1 whereas τ_{42} follows σ_4 and ρ_2.

Assuming that agents are senders half of the time, the expected utility of τ_i interacting with τ_j is defined by:

$$\text{EU}(\tau_i, \tau_j) = {}^1\!/\!_2\, \text{EU}_\sigma(\tau_i, \tau_j) \ + \ {}^1\!/\!_2\, \text{EU}_\rho(\tau_i, \tau_j). \tag{2.8}$$

The expected utilities of the 16 types of the symmetrized 2-states/messages signaling game are given in Table 2.5.

Finally, *fitness* indicates how well a type communicates in a population. Letting x be a population vector with x_j corresponding to the proportion of τ_j in x, the fitness of τ_i is defined as the average expected communicative success of this type given the type frequencies of the current population x:

$$f_i = \sum_j x_j \ \text{EU}(\tau_i, \tau_j) \tag{2.9}$$

Evolutionary stable strategies. Returning to the question about a population-level counterpart to a static solution concept such as that of a Nash equilibrium, a type's stability is classically associated with the notion of invasibility (Maynard Smith and Price 1973). Intuitively, an evolutionary stable strategy (ESS) is a strategy such that, if a population consist mostly of it, then this population is resistant against invader types changing the population's composition. More simply put: a type is evolutionary stable if its population cannot be taken over by others. This concept, just as that of a Nash equilibrium, is static: it does not make explicit which processes may lead a different type to join or take over the population. Similarly, it does not appeal to the rationality of members of a population; nor their knowledge of other types behavior; nor to their knowledge of the structure of the game. Instead, types are compared solely based on their (expected) utility following the intuition that a strategy is an ESS if fitness-based selection is sufficient to drive out invaders.

The concept of an ESS is a refinement of that of a Nash equilibrium. Consider a population consisting solely of τ_{11}. While $\langle \tau_{11}, \tau_{11} \rangle$ is a Nash equilibrium, such a population is tolerant toward any proportion of the 15 remaining types being present in it. As before, inspect the row of τ_{11} in Table 2.5 to see this. For instance, if an agent of type τ_{22} entered this population it would do as well as τ_{11}-players. What is more, not only would τ_{22} not be driven out of the population, if more players of this type emerged then τ_{22} would do better than τ_{11}. This would allow it to invade this initially monomorphic population. In short, Nash equilibria are too permissive to qualify as ESS. A possible solution would be to instead identify only strict Nash equilibria with stability at a population level. This move turns out to be too restrictive. Many games have mixed Nash equilibria that should count as stable (at the individual level; probabilistic strategies), but mixed Nash equilibria cannot be strict. In light of these considerations, Maynard Smith and Price (1973) propose that τ_i is an ESS iff

$$\mathrm{EU}(\tau_i, \tau_i) \geq \mathrm{EU}(\tau_j, \tau_i) \text{ for all alternative types } j; \tag{2.10}$$

$$\mathrm{EU}(\tau_i, \tau_i) = \mathrm{EU}(\tau_j, \tau_i) \rightarrow \mathrm{EU}(\tau_i, \tau_j) > \mathrm{EU}(\tau_j, \tau_j). \tag{2.11}$$

In words, τ_i is an ESS if, according to (2.10), it fares as least as well when interacting with itself than does any other type. Additionally, according to (2.11), should there be a type that fares as well against τ_i as τ_i against itself, then that type fares worse against its own type. The relationship between Nash equilibria and ESS can be summarized as follows (Nowak 2006):

$$\text{strict Nash equilibrium} \Rightarrow \text{ESS} \Rightarrow \text{weak Nash equilibrium}$$

A second common take on ESS is to make the degree to which an evolutionary outcome is impervious to invasion explicit, using an invasibility threshold $\bar{\epsilon}$. In this case, τ_i is an ESS iff for all $\epsilon < \bar{\epsilon}$ and all alternative types j:

$$\epsilon \mathrm{EU}(\tau_i, \tau_j) + (1-\epsilon)\mathrm{EU}(\tau_i, \tau_i) \; > \; \epsilon \mathrm{EU}(\tau_j, \tau_j) + (1-\epsilon)\mathrm{EU}(\tau_j, \tau_i). \tag{2.12}$$

That is, if the proportion of invaders is below threshold $\bar{\epsilon}$ then a population consisting of a proportion of $(1-\epsilon)$ types τ_i cannot be taken over. The underlying intuition of this definition remains the same as that of conditions (2.10) and (2.11).

2.3 On the Use and Limits of Static Equilibrium Analysis

Nash equilibria and ESS are informative about optimal outcomes at the individual and population level, respectively. The agnosticism of either type of static equilibrium analysis with respect to rationality and underlying processes is part of its appeal. Were it demonstrable that static predictions coincided with dynamic ones under fairly general conditions, our task would be restricted to that of finding a suitable model for a linguistic phenomenon under scrutiny and to then identify an appropriate static solution concept. There are, however, three major issues that static equilibrium methodology faces at either level. The first two are technical (Huttegger and Zollman 2013); the third is conceptual (Franke 2013).

Unclear predictions. First, as illustrated by the simple 2-states/messages case above, many signaling games have multiple equilibria. As mentioned earlier, Lewis' (1969) setup in fact guarantees a multiplicity of strict Nash equilibria at the individual level. The existence of multiple equilibria percolates to the population level in these cases as well, as only these strict Nash equilibria are ESSs (Wärneryd 1993). Prima facie, static equilibrium analysis does not offer guidance as to what its predictions are when more than one solution exist. This leaves such an analysis at best incomplete.

Uncertain predictions. With Searcy and Nowicki (2005) one may nevertheless intuit that optimal signaling is expected to emerge in Lewisian signaling games because agents have a common interest in information transfer. Any equilibrium would do, and there might appear to be little mystery to the fact that one of them will emerge. If so, a dynamic analysis would then "only" inform us about the process that leads to the adoption of one equilibrium over the other (see the third issue discussed below on this matter). However, a second problem of static equilibrium analysis is that neither at the individual nor at the population level these outcomes necessarily obtain.

Two simple and well-studied processes that illustrate this issue are Roth-Erev reinforcement learning, at the individual level, and the replicator dynamic, at the population level. These processes are particularly relevant to this issue because both, in principle, promote high utility behavior. They may therefore be expected to favor the optimal outcomes predicted by static equilibrium analysis.

Roth-Erev reinforcement learning is a simple learning process whereby actions that were successful in the past are rendered more likely to be chosen in the future (Roth and Erev 1995, Erev and Roth 1998). In the context of signaling games this means that a successful interaction will lead to a stronger association of the conveyed state with the used message in the case of senders, and to a stronger

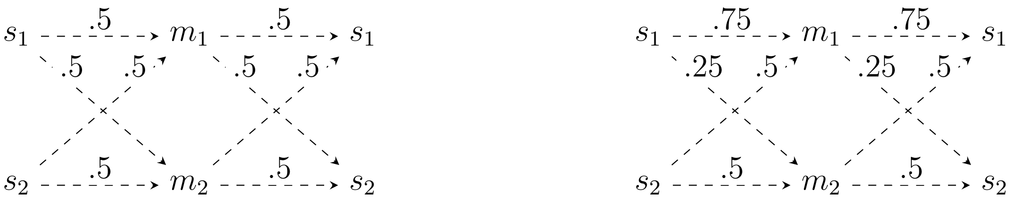


Figure 2.3: Roth-Erev reinforcement learning with initial weights set to 0.5 and r equal to player utility in a cooperative cheap talk game. Edge values show sender and receiver choice probabilities prior to any interaction (left), and after a first successful interaction conveying s_1 using m_1 (right).

association of the received message with the inferred state in the case of receivers. Prior to any interaction, each agent's associations are initialized with a weight. After an interaction relevant weights are then modified by a reinforcement value r. In each interaction the probability of choosing an action – to send a message or to interpret a message as a state – is proportional to its weight. Figure 2.3 illustrates this process. Initially, choice is driven by chance or whatever information fed the initial weight of a state-message or message-state association. However, over time, actions that were previously successful are more likely to be taken.

The replicator dynamic is one of the most famous population dynamics in evolutionary game theory. It models fitness-based selection, where the relative frequency of a type in a population increases with a gradient proportional to its average fitness in the population (Taylor and Jonker 1978, Hofbauer and Sigmund 2003; see Chapter 4 for details). This dynamic is popular and versatile because it can be derived from many abstract processes of biological and cultural transmission and selection (for overview and several derivations see Sandholm 2010), including conditional imitation (e.g., Helbing 1996, Schlag 1998) or reinforcement learning (e.g., Börgers and Sarin 1997, Beggs 2005).

Figure 2.4 shows the predictions of Roth-Erev reinforcement learning and the replicator dynamic for the cheap talk 2-states/messages signaling game. As showcased, neither dynamic guarantees the outcome predicted by its respective static equilibrium analysis. That is, not all dyads/populations converge to signaling systems. Instead, the proportion of suboptimal outcomes increases with the difference in frequency between the two states (see Catteeuw and Manderick 2014 for detailed analysis of reinforcement learning in signaling games, and Huttegger 2007 and Pawlowitsch 2008 on the replicator dynamic).

Many variants of these processes have been studied in connection to signaling games. For instance, reinforcement may not only be positive but punish actions that led to failure (e.g., Roth and Erev 1995). Alternatively, the stronger association of a state with a message may decrease this state's association with other messages (e.g., Franke and Jäger 2011). The replicator dynamic can similarly be supplemented by perturbations in the form of type mutations at generational

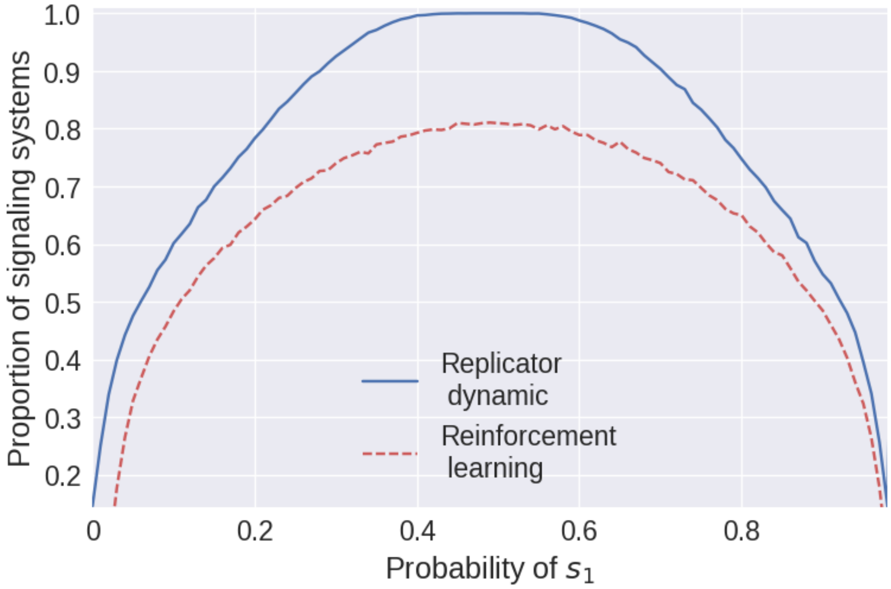


Figure 2.4: Proportion of signaling systems established under Roth-Erev reinforcement learning and the replicator dynamic in a cheap talk 2-states/messages signaling game across state probabilities $P^* \in \Delta(S)$. Reinforcement results correspond, for each probability of s_1, to the mean outcome of 20000 independent games of 500 interactions each with initial weights set to 1 for all associations and r equal to players' utility. Dyads with an expected utility higher than .95 after 500 interactions were categorized as having established a signaling system. Replication results correspond to mean proportions of 20000 independent populations after 500 generations, each randomly initialized.

turn-overs; or by dropping the assumption that players are equally likely to encounter each other, instead correlating encounters with expected utility or an environment's topological space (see, e.g., Skyrms 2010:§5). Some of these variations are more likely to converge to optimal outcomes than others. The crucial point, however, is that making these dynamics explicit is necessary to understand under which conditions an outcome can be expected to obtain. As summarized by Skyrms (2010:72): "The emergence of a signaling system is not always a moral certainty."

Explanatory force. The third problem of static equilibrium analysis has to do with the quality of its explanatory potential. As argued by Franke (2013) appeals to optimality are unsatisfying if the conditions under which an outcome emerges are not made explicit, or if the assumptions underlying a solution concept lack justification. Put differently, even if Searcy and Nowicki's (2005) intuition were correct, we would arguably not have learned much about the emergence and stability of a communication system. Instead, an explanatory analysis should be able to answer questions such as how rational agents are required to be, what information they base their choices on, what adaptive mechanisms they (minimally) need to possess, and so on. By virtue of its procedural agnosticism, static analysis cannot answer these questions.

With these issues in mind, we should stress that static solution concepts are nevertheless useful to characterize outcomes with respect to measures of interest. For example, communicative success. Outcomes identified in this manner can then be critically compared to those borne out under particular processes, as well as to those that are evidenced empirically. Under this view, static equilibrium analysis is supplementary to dynamic analysis (Huttegger and Zollman 2013).

2.4 Individual-Level Behavior: Rational Language Use

Iterated interactions and population dynamics build on individual-level behavior. At this level dynamic alternatives to static solution concepts also exist, e.g., rationalizability (Bernheim 1984, Pearce 1984), Benz and van Rooij's (2007) Optimal Assertions model, the Rational Speech Act model (Frank and Goodman 2012, Goodman and Stuhlmüller 2013), and the iterated X-response family: iterated Best-, Cautious-, and Quantal-response (Jäger 2007b, Franke 2009, Franke and Jäger 2014). Common to these concepts is that they model players' choices in a single interaction as resulting from a reasoning process over beliefs about other players' behavior. This contrasts with the preceding exposition where individuals were assumed to follow a strategy without specifying why.

The idea that linguistic choice results from a process of mutual reasoning about rational language use brings us back to the Grice. However, there is an important difference between Grice's approach and the one pursued here, which we briefly touched upon earlier. According to Grice (1975; 1989) the process of reasoning that interlocutors engage in is guided by the mutual assumption that certain principles of (rational) language use are followed; e.g., to be succinct, orderly, and relevant (see Chapter 1). Most dynamic game-theoretic approaches to pragmatics also view mutual reasoning as the motor that leads interlocutors to their choices. However, they do not rely on the formulation of conversational principles to constrain and guide communication. Instead, these approaches ground linguistic choice in individuals' contextual beliefs and preferences. Under this view, the explanation of linguistic behavior is not in term of maxims but in terms of rational behavior according to such beliefs and preferences (Benz 2006, Benz and van Rooij 2007, Franke and Jäger 2016b:120f). In informal terms, in cooperative communication interlocutors do their best to ensure faithful information transfer according to their beliefs about others' linguistic behavior and their own communicative preferences. Not because they believe that a particular rule of conversation is being followed. In virtue of not being tied to explicit conversational rules, such an approach also allows for predictions in situations not covered by Grice's principles, such as non-cooperative communication (e.g., De Jaegher and van Rooij 2014, Ahern and Clark 2017).

In recent years the idea that pragmatic reasoning can be captured by models that characterize linguistic inference as a product of explicit representations of mutual reasoning about beliefs and preferences has led to a diverse and growing literature. On a general level, this approach to pragmatics as *rational language use* encompasses game-theoretic approaches (e.g., Parikh 1991; 2000, Benz 2006, van Rooij and Sevenster 2006, Benz and van Rooij 2007, Jäger 2007b, Franke 2013, De Jaegher and van Rooij 2014, Franke and Jäger 2014) as well as Bayesian approaches (e.g., Frank and Goodman 2012, Goodman and Stuhlmüller 2013, Franke and Degen 2016, Bergen et al. 2016, Goodman and Frank 2016). Franke and Jäger (2016a:§3) identify five central properties common to these approaches: They are probabilistic, interactive, rationalistic, computational, and data oriented. The first four properties should not come as a surprise in light of the preceding discussion. The fifth refers to the fact that these models are usually not constrained to categorical predictions but allow for finer-grained quantitative predictions. This enables for a closer fit between theoretical predictions and actual communicative behavior evidenced, e.g., in experiments with human participants (Franke and Jäger 2016a:14). We will take advantage of this ability in Chapter 3. The remainder of this chapter incrementally introduces the common elements that make up models of rational language use.

2.4.1 Focal points and semantic meaning

For information transfer to take place in single interactions, the reasoning process that sender and receiver engage in should lead them to behave in a congruent fashion. That is, if things go smoothly, the receiver's interpretation of a message should agree with the information state the sender is attempting to convey. However, only a belief in common rationality is too weak to ensure meaningful predictions in one-shot signaling games (see Franke 2009:§1.2 for details on this negative result). Intuitively, the problem is that many signaling games do not have a unique "obvious" solution that sender and receiver can reach independently, without prior communication and only through mutual reasoning. Put differently, these games lack a solution that is noteworthy relative to others in that it would allow all interlocutors to conclude the others' reasoning process to be drawn to it. This issue was already touched upon from a different angle in Section 2.3, where we saw that Nash equilibria do not fulfill the requirement of uniqueness which could otherwise make them stick out relative to other strategies. To get meaningful inference without prior interaction off the ground the relation between the set of information states S and the set of messages M needs to be constrained.

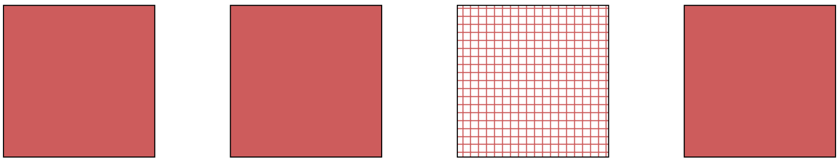


Figure 2.5: Illustration of a coordination problem without prior communication.

Focal points

Imagine that you and your partner are independently shown the four squares depicted in Figure 2.5 and that you have no means to communicate with each other. You are told that you will win a prize if both of you pick the same square. Which square would you choose? Similarly, imagine that you get lost in Amsterdam's city center and cannot get in touch with your partner. Where would you attempt to meet?

At a fundamental level these problems are analogous to the coordination problem posed by signaling games: You and your partner care about performing congruent actions but in principle any behavior/square/location could be chosen. Crucially and differently from the games discussed so far, in the coordination problems sketched above some squares/locations will appear to be more obvious candidate solutions than others. This may seem particularly evident for the problem illustrated in Figure 2.5. It is therefore important to stress that, in terms of the problem's setup, there is no reason to pick one square over the other. Coordination on a particular square is payoff irrelevant. You win the prize if you agree on any of them. However, things are different because you may believe that the third square draws you partner's attention and the belief that she believes it draws yours as well is a good reason to choose it. Such beliefs need not but may well be partner specific. For instance, while Amsterdam's Dam Square may be a more obvious location to rendezvous for residents, Amsterdam central station may be more salient to tourists.

Coordination problems such as these are what Schelling (1960) uses to illustrate the idea of *focal points*: solutions that are prominent, conspicuous or salient. In virtue of drawing reasoners' attention such solutions can improve coordination in the absence of prior agreement and lack of utility-relative differentiation (see, e.g., Mehta et al. 1994 for behavioral experiments supporting this claim). As conceived by Schelling, focal points come into play to break a tie between strategies. They are consequently of particular relevance when the game itself lacks such a tie breaker. For our purposes, the question is then whether within communication there is a plausible constraint on the beliefs interlocutors' may entertain about each other's linguistic behavior. As argued by Franke (2009:§2.1.1), semantic meaning is a natural candidate to serve this purpose.

Semantic meaning

The idea that semantic meaning plays a role in constraining what messages are admissible in which information state, although not necessarily its conception as a focal point, has figured in many game-theoretic approaches to rational language use (e.g., Parikh 1992; 2000, Benz 2006, Jäger 2007b, Stalnaker 2006, Benz and van Rooij 2007) and bleeds into classic literature on message credibility in signaling games (e.g., Rabin 1990, Matthews et al. 1991, Farrell 1993). In informal terms, the problem of message credibility is that, prima facie, the receiver recognizing that the sender intends her to believe the meaning of a message does not need to imply that she will form this belief (Stalnaker 2006). A message is credible in case the receiver can reason the sender to send it only in case it is true. Put differently, message credibility concerns the question whether there are situations in which the sender has incentive to be untruthful. As mentioned earlier when discussing preferences and cooperation, this problem is pressing if there is at least some conflict of interest between interlocutors. If instead the game is (believed to be) one of pure coordination there is no reason to suspect foul play.

Message credibility is particularly important in the context of fully rational agents and solution concepts, particularly epistemic ones, for it answers the question whether reasoning stays within the boundaries of truth. As already foreshadowed in Section 1.2, our analysis will not be concerned with these matters. Nevertheless, in line with this literature, semantic conventions and some degree of rationality will play a crucial role in providing the necessary fuel for meaningful pragmatic inference. In all generality, giving up the assumption of full rationality and allowing agents to occasionally err does however mean that we have to give up the claim that messages will always be used truthfully. In practice, the degrees of rationality we assume, while far from full rationality, will be sufficient to ensure that there is a strong tendency toward truthfulness in the games analyzed.

Returning to the conception of semantic meaning as a focal point, this view ascribes semantics a similar role to that of precedence and salience: it biases an agent's decision procedure. As illustrated below, such a bias proves to be sufficiently restrictive to reign in the multiplicity of inferences otherwise obtained, while providing sufficient flexibility as to not necessarily determine the outcome of the reasoning process on which it is based. Different from Schelling's (1960) conception of a focal point as a concept that applies at the final stage of deliberation, semantic meaning serves as a constraining point of departure for pragmatic reasoning. In other words, semantics provides the initial focal points that guide and constrain pragmatic inference.

In models of rational language use, semantic meaning is standardly represented by a lexicon L which maps state-message pairs to the (Boolean) truth-value of the message in that state. As a model's object, a lexicon represents the relevant semantic information required for pragmatic reasoning to get off the ground, relative to the phenomenon at hand. As illustrated shortly, the addition of individuals' beliefs and preferences to the model then yields the minimal set of relevant context features for pragmatic reasoning. A full specification of a game's lexicon together with the sender's and receiver's cognitive make-up can accordingly be construed as a *context model* (Franke and Jäger 2014:§1) for a class of pragmatic inferences (e.g., scalar implicatures). A convenient way to represent lexica is by $(|S|, |M|)$-Boolean matrices.

2.4.2 Pragmatic reasoning

Models of rational language use have been applied to many linguistic phenomena, ranging from implicatures (e.g., Benz and van Rooij 2007, Jäger 2007b, Franke 2009, Frank and Goodman 2012, Franke and Jäger 2014, Bergen et al. 2016) and disambiguation (e.g., Parikh 2000) to the use of generics (Tessler and Goodman 2016), polite language use (Yoon et al. 2016) and prosodic emphasis (Bergen and Goodman 2015), to name a few. To better showcase commonalities and differences across models, we will focus on one prominent kind of pragmatic inference that has received substantive attention in the literature: scalar implicatures.

Scalar implicatures

Scalar implicatures are a particularly productive and well studied class of systematic pragmatic inferences (Horn 1984, Hirschberg 1985, Levinson 1983, Geurts 2010). Usually, the utterance of a sentence like *I own some of Johnny Cash's albums* will be taken to mean that the speaker does not own all of them. This is because, if the speaker instead owned them all, she could have used the word *all* instead of *some* in her utterance, thereby making a more informative statement. Weak scalar expressions such as *some* are often semantically characterized as having literal meanings that are compatible with that of more informative relevant alternatives, like *all.* That is, if it were true that I own some of the albums, the literal meaning of *some* would not rule out that I own all of them. However, the use of a less informative expression when a more informative one could have been used can license a pragmatic inference that the stronger alternative does not hold. This rationale has been proposed to underlie the pragmatic use of many so-called scalar expressions. As exemplified by (2a) – (4a) and the defeasible scalar inferences they can give rise to in (2b) – (4b), English examples include numerals such as *five* and *six*, scalar adjectives like *warm* and *big*, as well as other quantifiers like *many.*

(2) a. I own some/many of Johnny Cash's albums.
b. $\rightsquigarrow$ I do not own all of Johnny Cash's albums.

(3) a. I have five dogs.
b. $\rightsquigarrow$ I do not have six/seven/... dogs.

(4) a. The soup is warm.

b. ⇝ The soup is not hot.

Scalar implicatures, especially the inference from *some* to *some but not all*, have been studied extensively, both theoretically (e.g., Horn 1984, Sauerland 2004, van Rooij and Schulz 2004, Chierchia et al. 2012) as well as experimentally (e.g., Bott and Noveck 2004, Huang and Snedeker 2009, Grodner et al. 2010, Goodman and Stuhlmüller 2013, Degen and Tanenhaus 2015). While there is much dispute in this domain about many details, a clear majority endorses the view that a weak scalar item like *some* is underspecified to semantically mean *some and maybe all* and that the enrichment to *some but not all* is part of some regular process with roots in pragmatics (see Chapter 4 for further discussion and analysis).

A minimal game-theoretic rendering of the semantics relevant for scalar implicatures is as follows: Assume that there are two relevant world states $S = \{s_{\exists\neg\forall}, s_\forall\}$. In state $s_{\exists\neg\forall}$ Chris owns some but not all of Johnny Cash's albums while in $s_\forall$ Chris owns them all. Additionally, assume that there are two relevant messages $M = \{m_{\text{some}}, m_{\text{all}}\}$, where m_{some} is short for a sentence like *Chris owns some of Johnny Cash's albums* and m_{all} is short for the same sentence with *some* replaced by *all.* Lexica for this case would assign a Boolean truth-value, either 0 for false or 1 for true, to each state-message pair. The majority view of semantically underspecified *some* is captured by the following lexicon:

$$L = \begin{array}{c} \\ s_{\exists\neg\forall} \\ s_\forall \end{array} \begin{array}{c} \begin{array}{cc} m_{\text{some}} & m_{\text{all}} \end{array} \\ \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \end{array}$$

In words, according to L the message m_{some} is true of both $s_{\exists\neg\forall}$ and $s_\forall$ whereas message m_{all} is true only of state $s_\forall$.

Rational language use

What behavior would a rational user of L exhibit? A rational hearer would reason about the message she receives from the speaker, taking as a point of departure the semantics of L. Intuitively and in analogy to the characterization of scalar implicatures given above, if a rational speaker means to convey $s_\forall$ she should send m_{all}. This message is semantically unequivocal, thereby increasing the chance of communicative success. In the case of $s_{\exists\neg\forall}$, semantically, only message m_{some} is an option, but this message could signal either state. Nevertheless, if m_{all} is reasoned to signal $s_\forall$, then the hearer can infer that m_{some} is to be associated with $s_{\exists\neg\forall}$ not with $s_\forall$. A rational speaker reasons in analogous fashion, coming to her behavior through reasoning about the hearer's likely interpretation of a message. She will accordingly send m_{some} in $s_{\exists\neg\forall}$ and m_{all} in $s_\forall$.

Models of rational language use such as the Optimal Assertions model (Benz 2006, Benz and van Rooij 2007), the Rational Speech Act model (Frank and Goodman 2012, Goodman and Stuhlmüller 2013) and Iterated Best-, Cautious- and Quantal-Response models (Jäger 2007b, Franke 2009, Franke and Jäger 2014) represent this reasoning procedure as a hierarchy over reasoning types. The bottom of the hierarchy, level 0, corresponds to literal signaling behavior. Literal language users do not reason about their interlocutors but simply produce/comprehend according to their preferences/expectations and the semantics provided by their lexica. Player i's literal receiver and sender behavior are defined in (2.13) and (2.14).

$$\rho_0(s|m; pr^i; L) \propto L_{[s,m]}\, pr^i(s); \tag{2.13}$$

$$\sigma_0(m|s; L) \propto L_{[s,m]} - c^i_\sigma(m), \tag{2.14}$$

where pr^i is player i's prior over states, $pr^i \in \Delta(S)$. Such a prior represents the subjective relative saliency of states in a context. They are an individual's expectations in a particular situation. We construe such expectations as based on any source of information beyond an expression's literal meaning. Among others, a prior may draw from the context in which communication takes place, general expectations of language use, or perceptual information. In short, it represents condensed information of the association strength with which an interpretation comes to mind (Franke 2009:129ff). Whenever cost, priors, or lexica are assumed to be common – i.e., shared across individuals – we omit individual indices as well as their explicit codification as conditional parameters in choice probabilities.

To the best of my knowledge, in the case of scalar implicatures no strong evidence about differences in preferences between relevant message alternatives has been found, for example, between the choice of *some* and *all*; nor have particular prior biases been suggested. We may therefore assume that the scalar inference game is a cheap talk game with an uninformative common prior, $pr(s_\forall) = {}^1\!/\!_2 = pr(s_{\exists\neg\forall})$. These assumptions are further motivated by the broader question whether some pragmatic inferences can be explained purely in terms of mutual reasoning. That is, whether certain inference patterns can be explained without appeal to state saliency or differential message preferences. This should not be taken to suggest that cost or priors should not play an explanatory role in pragmatics. Nor that an explanation that appeals to these factors is subordinate to one that does not. Rather, in the face of a lack of evidence to the contrary, an explanation that does not *require* a particular prior or cost-assignment is more parsimonious than one that does not.

Under these assumptions and with L as above, definitions (2.13) and (2.14) give the following choice probabilities for literal behavior in the some-all context model:

$$\sigma_0(\cdot \mid \cdot) = \begin{array}{c} \\ s_{\exists\neg\forall} \\ s_\forall \end{array} \begin{array}{c} \begin{array}{cc} m_{\text{some}} & m_{\text{all}} \end{array} \\ \begin{bmatrix} 1 & 0 \\ .5 & .5 \end{bmatrix} \end{array} \qquad \rho_0(\cdot \mid \cdot) = \begin{array}{c} \\ m_{\text{some}} \\ m_{\text{all}} \end{array} \begin{array}{c} \begin{array}{cc} s_{\exists\neg\forall} & s_\forall \end{array} \\ \begin{bmatrix} .5 & .5 \\ 0 & 1 \end{bmatrix} \end{array}$$

In words, a literal sender with no preferences over messages will be indifferent between sending m_{some} or m_{all} for $s_\forall$, sending either with equal probability, and she will send m_{some} to convey $s_{\exists\neg\forall}$. Analogously, if their prior is flat, literal interpreters choose an arbitrary true interpretation for each message according to their lexicon.

Most models of rational pragmatic language use assume that naïve level-0 reasoners do not correspond to actual linguistic behavior. This is reflected by the fact that neither choice rule involves (an approximation of) utility maximization. Instead, (2.13) and (2.14) characterize completely unreflected literal language use.

Higher order reasoning types of level $n + 1$ make their linguistic choices according to the expected behavior of a level n interlocutor. There are two general approaches to how higher order reasoning is set up. The first assumes that agents are strictly rational utility maximizers (e.g., Benz and van Rooij 2007, Jäger 2007b, Franke 2009). The second approach allows for some slack in agents' tendency toward utility maximization (e.g., Goodman and Stuhlmüller 2013, Franke and Jäger 2014). As motivated below, we will follow the latter approach. For a finite state space this behavior is defined as follows for player i:

$$\rho_{n+1}(s|m; pr^i; L) \propto \exp(\lambda \frac{\sigma_n(m|s; L)\; pr^i(s)}{\sum_{s'} \sigma_n(m|s'; L) pr^i(s')}); \tag{2.15}$$

$$\sigma_{n+1}(m|s; L) \propto \exp(\lambda(\rho_n(s|m; pr^i; L) - c_\sigma(m))). \tag{2.16}$$

That is, instead of using only the literal meaning of messages, higher order reasoners use Bayes' rule to weigh their possible actions based on a conjecture about their interlocutor's behavior. In both cases choice is regulated by a soft-maximization parameter $\lambda \geq 0$ (Luce 1959, Sutton and Barto 1998). As λ increases choices approach strict maximization of expected utilities. For a sender this means that messages reasoned to have a high probability of being understood that are of low cost are increasingly prioritized over low success and/or high cost ones. In the case of receivers, states consistent with a conjecture of rational speaker behavior that are favored by their prior over states are more likely to be inferred. This so-called *rationality parameter* thereby allows for the representation of a range of behavioral strategies: from irrational behavior ($\lambda = 0$) up to the approximation of strictly rational behavior as λ approaches infinity.

In contrast to strict utility maximization, soft-maximization has the advantage of making explicit the degree to which rational behavior is necessary to explain pragmatic inference. Additionally, it allows us to consider cases in which some deviation from optimal rational behavior might even be explanatory. From a more technical perspective, soft-maximization also has the desirable consequence that no choice is ever completely ruled out – even if acting upon it may be highly unlikely. By contrast, models that assume strict utility maximization require an additional assumption to deal with situations in which receivers reason that a message will never be sent (see Franke 2009, Franke and Jäger 2014, and Bergen et al. 2016:2ff for details and discussion). Such cases require attention because receiver strategies need to specify the behavior that ensues after the reception of any message. Otherwise, if, for one reason or another, such a surprise message is used, the receiver's reaction to it would not be defined.

Another desirable consequence of not having degenerate choice probabilities will become relevant once iterated Bayesian learning comes into the picture (see Chapter 4 for details). That is, once we consider cases in which agents need not just use but first acquire language. In such situations non-zero choice probabilities allow naïve learners to entertain that all learning hypotheses could be compatible with the input they receive. Degenerate counterparts could, by contrast, lead to situations in which a language is faithfully transmitted solely because it harbors a single word/feature/construction that no user of a different language would ever use in virtue of strict utility maximization. Put differently, (at least minimal) variation in linguistic behavior opens up to the possibility of variation in the transmission of linguistic knowledge across generations.

Returning to single interactions and linguistic choice, as specified by (2.15) and (2.16), higher order reasoning types are assumed to behave rationally according to the expected behavior of a level n interlocutor. This is a debatable design choice. For instance, a more flexible – and possibly more realistic – alternative would be for players to have beliefs about their interlocutor's level of sophistication and for choice probabilities to be derived from these beliefs (see, e.g., Camerer et al. 2004). The assumption that players believe their interlocutor to be exactly one level less sophisticated than themselves is first and foremost made for simplicity. It primarily hinges on a trade-off between a more rigid reasoning procedure and increased model complexity. However, this assumption of myopic reasoning types has also been shown to succeed in predicting various empirically attested linguistic patterns (see Goodman and Frank 2016 for a recent overview). We will therefore opt for this simpler, and to my mind more perspicuous assumption, while bounding agents' reasoning to a low degree of sophistication: level 1. This minimal degree of mutual reasoning will suffice to capture the classes of inferences we aim to characterize in this investigation.

To illustrate how higher order reasoning affects linguistic behavior, let us turn to the some-all context model again. With $\lambda = 1$ the choice probabilities of level-1 players are as follows:

$$\sigma_1(\cdot \mid \cdot) \approx \begin{array}{c} \\ s_{\exists\neg\forall} \\ s_{\forall} \end{array} \begin{array}{c} \begin{array}{cc} m_{\text{some}} & m_{\text{all}} \end{array} \\ \begin{bmatrix} .62 & .38 \\ .38 & .62 \end{bmatrix} \end{array} \qquad \rho_1(\cdot \mid \cdot) \approx \begin{array}{c} \\ m_{\text{some}} \\ m_{\text{all}} \end{array} \begin{array}{c} \begin{array}{cc} s_{\exists\neg\forall} & s_{\forall} \end{array} \\ \begin{bmatrix} .58 & .42 \\ .27 & .73 \end{bmatrix} \end{array}$$

By contrast, with $\lambda = 10$ we have:

$$\sigma_1(\cdot \mid \cdot) \approx \begin{array}{c} \\ s_{\exists\neg\forall} \\ s_{\forall} \end{array} \begin{array}{c} \begin{array}{cc} m_{\text{some}} & m_{\text{all}} \end{array} \\ \begin{bmatrix} .99 & .01 \\ .01 & .99 \end{bmatrix} \end{array} \qquad \rho_1(\cdot \mid \cdot) \approx \begin{array}{c} \\ m_{\text{some}} \\ m_{\text{all}} \end{array} \begin{array}{c} \begin{array}{cc} s_{\exists\neg\forall} & s_{\forall} \end{array} \\ \begin{bmatrix} .97 & .03 \\ 0 & 1 \end{bmatrix} \end{array}$$

The intuition behind the stronger association of m_{some} with $s_{\exists\neg\forall}$ and that of m_{all} with $s_{\forall}$ follows the same rationale given earlier. It results from agents' mutual reasoning about (boundedly) rational language use.

More generally, in the some-all context model the association of $s_{\exists\neg\forall}$ with m_{some} and that of $s_{\forall}$ with m_{all} approximates 1 with $\lambda \geq 17$ for level-1 reasoning. For level-2 reasoning this result already obtains for $\lambda \geq 7$. Overall, the pragmatic strengthening of an underspecified message, in this case m_{some} to signal *some but not all* rather than *some and maybe all*, is predicted provided (i) some degree of rationality in choice (high λ) and (ii) some degree of mutual reasoning (n-level reasoning with $n \geq 1$). Importantly, higher level reasoning can only lead to this strengthening if fueled by a common belief in rational language use. As suggested above by $\lambda = 1$, in cases of low rationality choice probabilities instead approach .5 as reasoning levels increase.

Variation across models

Models of rational language use share many fundamental components: they represent mutual reasoning as a reasoning hierarchy and characterize pragmatic inference as based on explicit representations of beliefs, preferences, and semantic conventions. The main difference across approaches lies in the choice functions of reasoning types and the depth of the reasoning on which they build. Four influential models that differ in this respect are the Rational Speech Act (RSA) model (Frank and Goodman 2012, Goodman and Stuhlmüller 2013), the Optimal Assertions (OA) model (Benz 2006, Benz and van Rooij 2007), the Iterated Best Response (IBR) model (Jäger 2007a, Franke 2009, Franke and Jäger 2014), and the Iterated Quantal Response (IQR) model (Franke and Jäger 2014).

In contrast to OA as well as to IBR and IQR, the RSA model was originally formulated as a hearer-centric model. Its main focus accordingly laid in capturing pragmatic interpretation rather than speaker behavior. The RSA model defines level-0 behavior and level-1 sender behavior in a similar fashion to definitions (2.13) and (2.16). However, in contrast to the above, no literal sender is defined. Instead, pragmatic interpretation is identified with a level-2 receiver who reasons about level-1 sender behavior. Put differently, early instances of the model focused exclusively on the reasoning chain that defines level-2 receiver behavior. There are two other major differences to our definitions. First, only level-1 senders are assumed to soft-maximize. A receiver's pragmatic inference instead only involves the inversion of level-1 speaker behavior using Bayes' rule. Second, utilities in RSA models are associated with an information-theoretic measure of a message's informativity about the sender state. For player i this gives the following RSA-style behaviors:[3]

$$\begin{aligned}
\rho_0(s|m;pr^i;L) &\propto L_{[s,m]}\, pr^i(s); && (2.17)\\
\sigma_1(m|s;L) &\propto \exp(\lambda(\log \rho_0(s|m;pr^i;L) - c_\sigma(m))); && (2.18)\\
\rho_2(s|m;pr^i;L) &\propto \sigma_1(m|s;L)\, pr^i(s). && (2.19)
\end{aligned}$$

In a nutshell, RSA views the sender's goal as that of inducing a belief about the state in the receiver (with minimal effort if message cost is involved) by sending informative messages. The receiver's goal is to form this belief.

Beyond numerical differences that can stem from the different ways in which level-1 sender behavior is defined, there are also conceptual differences between the RSA-approach and game-theoretical ones, embodied here by OA, RSA and IQR models. In the context of this investigation these differences are slightly obscured by the fact that we focus on interpretation games. To illustrate the two conceptions more clearly, recall that, in a general form, receiver strategies are defined as mappings from messages to (distributions over) acts. As mentioned earlier, acts can be interpretations; but they could equally well be reactions to animal alarm calls, such as hiding in a tree or hiding in a bush, or any other physical action such as passing the salt or bringing the keys. Under a game-theoretical view, utilities are preferences over state-message-act triples.

Following Qing and Franke (2015) we can call the RSA-view on utility *belief*-oriented and the game-theoretic view *action*-oriented. Under the former view, speakers care to maximize log $\rho_0(s \mid m) - c_\sigma(m)$. That is, first and foremost, they care about the receiver's beliefs and the receiver cares to form this belief. Under the latter view, receivers care to maximize their (expected) utility. It is their preference over state-(message-)act triples that matters.

The signaling behavior defined by OA, IBR, and IQR is more similar to definitions (2.13) – (2.16). Coming from a game-theoretic tradition, all three models are action-oriented as well. As for their differences, instead of soft-maximization, OA and IBR assume strict utility maximization at each step of the reasoning

[3] As with other models of rational language use, there is substantial variation within the RSA tradition. For instance, on whether receivers take senders' preferences over messages into consideration; or on whether receivers' priors only come into play at reasoning level 2, letting the prior of level-0 receivers instead be uniform across contexts. These finer differences do not need to concern us here, but see Qing and Franke (2015) for detailed comparison.

hierarchy. IQR assumes soft-maximization at every step. This contrasts with our definitions of literal behavior in (2.13) and (2.14), where no tendency toward utility maximization is assumed. IBR and OA differ in two respects (Franke 2008). First, OA limits the receiver's reasoning sequence to ρ_0-σ_1-ρ_2 in a similar fashion to RSA. Second, prior state probabilities do not come into play in the OA model. Of course, behaviorally this difference disappears if the prior is uniform.

These variations can be seen as reflecting the explanatory goal of each model. OA, IBR, and IQR come from a game-theoretic tradition and aim to capture the signaling behavior of both players to understand their joint outcome. OA and IBR follow the classic assumption that choice is strictly rational. IQR accommodates a more diverse range of behaviors in a trade-off between some analytic results of OA and IBR and the possibility to capture finer-grained predictions involving different degrees of rationality. This ability makes IQR models more attractive for investigations that involve empirical data (Degen and Franke 2012, Franke and Jäger 2016a; see Chapter 3 for a concrete application). In its original formulation, the RSA model is more constrained in that it focuses only on a particular kind of receiver behavior. This constraint is a consequence of its original goal to model experimental data involving only pragmatic interpretation. However, nothing prevents the model from being extended to characterize sender behavior as well (see, e.g., Franke and Degen 2016 for such an extension). In sum, I believe it is fair to say that the commonalities among these approaches far supersede their differences. Within any approach it is common to find a number of slight differences in how each analysis sets up the linguistic behavior of agents. However, these changes mainly hinge on the phenomenon to be explained. They do not purport to be fundamentally new proposals in their own right.

As for my own assumptions on these matters, the definitions in (2.13) – (2.16) and their instantiations in the following chapters are mainly motivated by the advantages conferred by soft-maximization in contrast to strict utility maximization mentioned above. As for other differences, contra IQR, level-0 soft-maximization is not assumed in cases where literal behavior is not considered to correspond to actual linguistic behavior.[4] Contra RSA, higher level receivers are assumed to soft-maximize because their goal is not only to form a posterior belief about the sender state, but rather to maximize the utility derived from the act they perform upon reception of a message. That is, I will follow the game-theoretic view of action-oriented signaling. On this matter we should note that whether there is a conceptual difference between belief- and action-oriented models depends on the situation at hand. In cases in which interlocutors care about belief-formation they evidently coincide (see van Rooy 2004b for detailed comparison of notions of utility and related information-theoretic measures). In terms of technical differ-

[4] In Chapter 4 we will come to contrast literal behavior with behavior that results from higher order reasoning under evolutionary dynamics. In this setup literal behavior is taken to correspond to behavior that can actually be adopted. Consequently we assume it to come with a rationality parameter that regulates choice (see Franke and Degen 2016 for a similar choice).

ences, the choice of one over the other is ultimately empirical. A first step toward their experimental comparison is provided by Qing and Franke (2015), who show that action-oriented behavior better reflected data collected in a reference game. Nevertheless, this issue is far from settled and invites future research.

2.5 Final Remarks

This chapter touched upon a number of different questions. The two central ones are:

(i) How can linguistic behavior be represented?

(ii) How can linguistic outcomes be analyzed?

On question (i) we follow the view of language use as a strategic endeavor of information transfer and employ signaling games as context models of features relevant to an interaction: individuals' preferences and beliefs, together with the linguistic material relevant to the situation at hand. At a general level, this view of communication identifies linguistic knowledge with operational knowledge (Parikh 1994:530ff). What is relevant for communication is not that speaker and receiver share a language. Rather, what is relevant is that their language use effects successful behavior. In the case of interpretation games this translates to an agreement between a sender's information state and a receiver's interpretation. Importantly, this view allows for mismatches in interlocutors' subjective priors (Chapter 3) or even their semantics (Chapter 4 and 5).

As for question (ii), we argued that, where possible, linguistic outcomes should be explained in terms of the processes that lead to their emergence and stability rather than by appeal to their optimality. Particular phenomena will of course call for the analysis of particular processes, at possibly distinct levels of interaction. Irrespective of these differences, the common thread of this investigation lies in that they build on models of rational language use at the individual level. This allow us to inspect the joint role of semantic conventions and language use in the emergence of pervasive phenomena at the semantics-pragmatics interface.

Chapter 3

Signaling Under Uncertainty

Interpretative Alignment Without a Common Prior

> To conclude, the light of human minds is perspicuous words, but by exact definitions first snuffed, and purged from ambiguity; reason is the pace; increase of science, the way; and the benefit of mankind, the end. And on the contrary, metaphors, and senseless and ambiguous words, are like ignes fatui; and reasoning upon them is wandering amongst innumerable absurdities; and their end, contention and sedition, or contempt.
>
> Thomas Hobbes, *Leviathan*

Communication involves a great deal of uncertainty. Prima facie, it is therefore surprising that biological communication systems – from cellular to human – exhibit a high degree of ambiguity and often leave its resolution to contextual cues. This puzzle deepens once we consider that contextual information may diverge among individuals.

In this chapter we lay out a model of iterated ambiguous communication between subjectively rational agents that lack a common contextual prior. On its basis, we argue ambiguity's justification to lie in endowing interlocutors with means to flexibly adapt language use to each other and the context of their interaction to best serve their communicative preferences. Linguistic alignment is shown to play an important role in this process. It foments convergence of contextual expectations and thereby leads to agreeing use and interpretation of ambiguous messages. We conclude that ambiguity is ecologically rational when (i) interlocutors' (beliefs about) contextual expectations are generally in line or (ii) they interact multiple times in an informative context, enabling for the alignment of their expectations. In light of these results, meaning multiplicity can be understood as an opportunistic device enabled and shaped by linguistic adaptation and contextual information.

3.1 Meaning Multiplicity in Communication

In principle, speakers can draw from a large and virtually inexhaustive pool of alternatives to convey a state of affairs. We can refer to an entity as *Donald Trump*, as *the forty-fifth president of the United States*, or simply as *that guy*. Similarly, we may say *bank* rather than *financial institute*, *bat* rather than *baseball club*, *superfluous hair remover* rather than *remover of superfluous hair*, or *thing* instead of any of the aforementioned. When the primary goal is information transfer the linguistic choices of speakers are chiefly constrained by whether their interlocutors will be able to infer information as intended. Why and when, then, would a speaker opt for a more ambiguous expression over one that is less ambiguous?

The diverse nature of the examples above illustrates the issue we seek to address in this chapter: from cellular signals to those employed by meerkats and baboons, biological signaling is rife with meaning multiplicity (Greenough et al. 1998, Arnold and Zuberbühler 2006, Santana 2014). Natural languages are no exception. Prima facie, this fact may be qualified as puzzling, if not as downright indicative for a lack of communicative efficiency in their design (Chomsky 2002; 2008).

Ambiguity avoidance has an intuitive appeal because the association of multiple meanings with a single expression can give rise to uncertainty in interpretation. Consequently, unambiguous language may be argued to be better suited for communication. This idea has prominently figured in investigations on the emergence of signaling systems, where an emerging system is standardly evaluated against the ideal of one-to-one form-meaning mappings (e.g., Lewis 1969, Steels 1998, Skyrms 2010; see Spike et al. 2016 for a recent review). Notwithstanding, a growing body of literature argues that meaning multiplicity confers functional advantages. It allows for smaller vocabularies (Santana 2014), greater signal compression (Juba et al. 2011), for the reuse of forms that are easier to produce or parse (Horn 1984, van Rooij and Sevenster 2006, Piantadosi et al. 2012b, Dautriche 2015), for the partition of large semantic spaces (O'Connor 2015), for coordination on non-lexicalized meaning (Brochhagen 2015b), and for deception in non-cooperative communication (Crawford and Sobel 1982). In most of these analyses the exploitation of contextual information plays a central role. The argument is simple: the information provided by context needs not be codified in a signal. As a consequence, ambiguous languages can be more compressed or enable for a more optimal reuse of their inventory than unambiguous counterparts, while transmitting information as faithfully.

A complication ignored by this justification is that the gain attained from contextual information is not necessarily cashed out in situations in which contextual information varies across agents. Once a divergence of subjective contextual information is admitted, it remains to be shown what the consequences of meaning multiplicity are in cooperative communication. Furthermore, even if the assumption of a common contextual prior were justified, it is not clear how it may come about nor how it relates to the context of interaction itself. In the following, we take up these challenges by analyzing ambiguous communication in iterated interactions without a common contextual prior. To this end, we propose a conservative generalization of the speaker behavior defined in (2.16) and then analyze its predictions when combined with simple adaptive dynamics. The result is a game-theoretic model that combines mutual reasoning with pragmatic uncertainty which allows players to adapt their language use to each other over time.

This chapter's main goal is twofold. First, we set out to investigate the conditions under which meaning multiplicity is advantageous by going beyond static approaches, as well as by decoupling context from the subjective access individuals may have to it. Second, we seek to further our understanding of the consequences of linguistic alignment by analyzing how the interplay of context, subjective contextual expectations, and iterated interactions shapes (un)ambiguous language use.

We proceed as follows. Section 3.2 discusses the issues that an analysis of ambiguity in terms of a common prior raises. Section 3.3 lays out our main assumptions together with the model we employ to characterize communication under pragmatic uncertainty. Section 3.4 showcases the model's main predictions and explores the consequences of possible refinements, as well as those that follow from environmental constraints. Going beyond theoretical predictions, Section 3.5 shows how well our model can explain experimental data. We critically assess our main findings and possible shortcomings in Section 3.6, and conclude in Section 3.7.

3.2 Ambiguity, Preferences, Context and Common Priors

In linguistics it is common to distinguish different types of meaning multiplicity based on syntactic, phonetic, graphemic, or semantic criteria. We instead take a decision-theoretic point of view under which the relevant distinction concerns whether or not communicative success hinges on the discrimination of interpretations conventionally associated with the same form (Parikh 2000:§6, Santana 2014:§3). That is, whether an expression requires its addressee to settle for a particular interpretation over others – be it simplex or complex, and irrespective of the locus of its meaning multiplicity. While this conception is broad, it excludes phenomena such as vagueness, where an expression may have multiple precifications but (at least partially) successful interpretation does not hinge in teasing them apart (see, for example, De Jaegher and van Rooij 2011, Franke et al. 2011, O'Connor 2014). For the sake of brevity we will call this property ambiguity, as tacitly done so far.

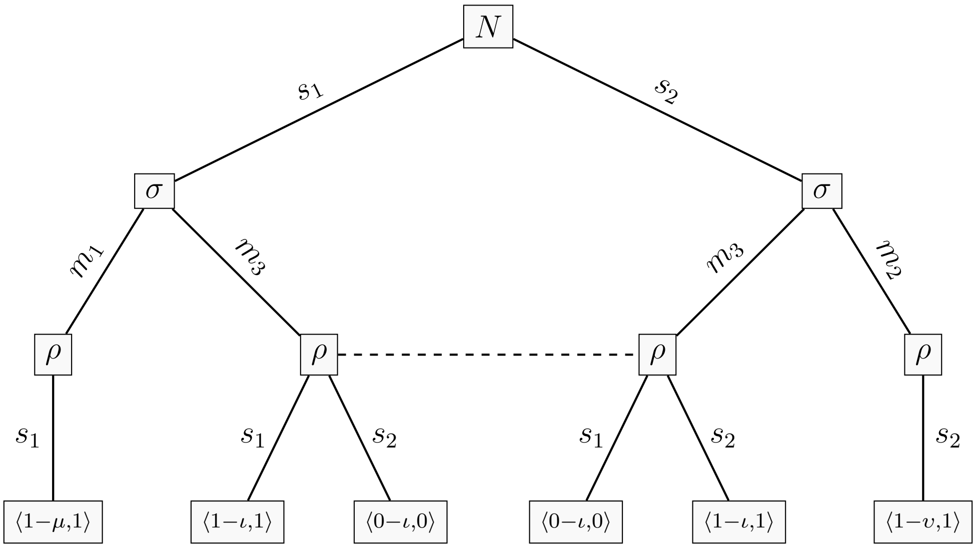


Figure 3.1: Sequential structure of Parikh's (2000) game. Starting at the top, nature selects a state, s_1 or s_2, then sender σ picks a (true) message to send in this state, and, lastly, receiver ρ interprets the message. End nodes indicate sender and receiver utility for a given branch, with μ being the cost associated with the use of m_1, υ that of m_2 and ι that of m_3. The dotted line represents the receiver's epistemic uncertainty about which branch she is in when receiving m_3.

As motivated in more detail below, we focus on situations where ambiguous signaling is deliberate insofar as the speaker could have chosen a less ambiguous expression. A minimal lexicon fragment that allows for the choice between ambiguous messages over unambiguous ones is one with three messages and two states, where $L(s_1, m_1) = 1 = L(s_2, m_2)$, $L(s_1, m_3) = 1 = L(s_2, m_3)$, and all other state-message pairings are false:

$$L = \begin{array}{c} \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{ccc} m_1 & m_2 & m_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{array}$$

In words, according to lexicon L message m_1 is exclusively true of state s_1, m_2 only of s_2 and m_3 is ambiguous between these two states. Speakers of L therefore need not use ambiguous m_3 to convey these states but may nevertheless choose to do so.

To better illustrate the technical and conceptual issues at stake, as well as to better motivate our own assumptions on these matters, we postpone our analysis of L in terms of a model of rational language use to Section 3.3. In this section, we instead use Parikh's (2000) analysis of ambiguity as a guide.

Figure 3.1 shows the sequential game that Parikh focuses on, with μ, υ and ι being the numerical cost the sender associates with messages m_1, m_2 and m_3, respectively (cf. Figure 4 in Parikh 2000:199). This game follows from the semantics specified by L if, with Parikh, we assume that sender and receiver behavior are fully constrained by semantic meaning. In Gricean terms, the assumption is that the maxim of quality – to be truthful – cannot be violated; or, using Benz and van Rooij's (2007) terminology, that only *admissible* messages are sent.

In this game, the sender's preferences over messages are inverse to the values of μ, ι and υ. If this were a cheap talk game, $\mu = \iota = \upsilon = 0$, then the strategic situation is clear: a rational sender should always send the unambiguous message with an index that matches her information state. This would leave no room for misunderstanding and ambiguity is avoided. In the face of ambiguity being a pervasive property of natural communication then either Chomsky (2002; 2008) is right and (the use of) m_3 indicates a defect of language (use), or ambiguity's justification lies in other factors such as a message bottleneck (O'Connor 2015) or non-cooperative communication (Crawford and Sobel 1982). However, as discussed below, there are reasons to believe that ambiguous signaling is often a deliberate choice in cooperative communication as well, and not necessarily driven by restrictions in the space of forms available. Such situations can be characterized through differential sender preferences over messages, where not all messages are equally preferred.

3.2.1 Brevity and context

While there can be many idiosyncratic reasons why a speaker may prefer a particular ambiguous message over a less ambiguous one, we will illustrate our predictions by assuming a speaker preference for brevity. As argued in the following, brevity is a plausible candidate for a preference shared across individuals. It additionally has a bearing on domains central to our purpose: linguistic choice, dialogal adaptation, ambiguity, and contextual predictability.

Brevity is often argued to be a rational speaker-oriented principle. It is posited, for instance, in Grice's (1975) maxim of manner, Horn's (1984) R-principle, and Zipf's (1949) principle of least effort. The tension between ambiguity and brevity is explicit in the interaction between Grice's maxims of quantity – to be as but not more informative than required – and his manner (sub)maxims – to be brief but to avoid ambiguity.

In dialog, message brevity has been reported to increase incrementally in iterated tasks (Clark and Wilkes-Gibbs 1986, Motamedi et al. 2016, Hawkins et al. 2017, Kanwal et al. 2017). This provides some indirect evidence for speakers seeking to increase message compression when possible.

In the case of word length, Piantadosi et al. (2011) report cross-linguistic evidence for its predictability based on contextual information; a prediction subsequently corroborated by Mahowald et al. (2013) in a behavioral study suggesting that this relation is a consequence of deliberate speaker choices, instead of a statistical effect of language use or word classes. That is, there is some supporting evidence for the assumption that brevity interacts with contextual information and influences linguistic behavior. The more predictive the context, the shorter messages tend to be. Furthermore, there is a wealth of evidence for a negative correlation between contextual predictability and the pronunciation length of phones and words (see Brennan and Hanna 2009:§2.1, Piantadosi et al. 2011, and references within).

The claim that ambiguity's risk is assessed through contextual rather than language internal factors has also received some empirical support (see Ferreira 2008 and Wasow 2015 for recent psycholinguistic overviews). Two main findings are relevant here. First, there is little evidence for the idea that ambiguity influences linguistic behavior to the extent that speakers always prefer unambiguous expressions over ambiguous ones. This is contrary to the idea that ambiguity avoidance exerts a strong influence on speakers' choices. Second, while no conclusive evidence for this kind of avoidance has been found, Ferreira et al. 2005 do report a tendency for the avoidance of ambiguity in naming tasks under certain conditions. This tendency was registered in situations in which a single reading of an ambiguous expression still applied to multiple objects. For example, subjects presented with multiple baseball bats of different sizes avoided the plain label *bat* to name one of them. The same degree of avoidance was not registered when the naming target was a baseball bat while a bat of the zoological kind was also present. A possible rationalization of this difference is that speakers may expect meaning multiplicity rooted in linguistic conventions to be manageable (*baseball bat* vs. *flying bat*) whereas more information is supplied when context typically would not lend sufficient support to a single interpretation (as is the case when a label applies to multiple objects of the same kind). As we argue in the following: for ambiguity to be advantageous, meanings associated to a single form should generally appear in contrasting contexts to safeguard understanding. That is, they should appear in contexts in which priors sufficiently favor one interpretation over the other. An expectation that the addressee will be able to resolve ambiguity may conversely not be warranted when the risk of misunderstanding stems from atypical or language external factors. Additionally, in Ferreira et al.'s task the choice of less ambiguous labels may have been fueled by the addressee being unknown to the speaker.

Returning to Parikh's game in Figure 3.1, the situation we are after is then one in which m_3 is shorter than m_1 and m_2, and thereby preferred: $c_\sigma(m_3) < c_\sigma(m_1)$ and $c_\sigma(m_3) < c_\sigma(m_2)$. Albeit sparsely motivated, this is the same preference ordering that Parikh (2000) assumes, with $c_\sigma(m_1) = c_\sigma(m_2)$.

3.2.2 Context and (beliefs about) subjective expectations

If the deliberate use of ambiguous expressions hinges on contextual factors, this begs questions about of the kind of contextual information interlocutors have access to. We will argue that to answer such questions satisfactorily, the conception

Sender strategies		Receiver strategies	
σ_1: $s_1 \mapsto m_1$ $s_2 \mapsto m_3$	σ_2 : $s_1 \mapsto m_3$ $s_2 \mapsto m_2$	ρ_1: $m_1 \mapsto s_1$ $m_2 \mapsto s_2$ $m_3 \mapsto s_2$	ρ_2 : $m_1 \mapsto s_1$ $m_2 \mapsto s_2$ $m_3 \mapsto s_1$
σ_3: $s_1 \mapsto m_3$ $s_2 \mapsto m_3$	σ_4 : $s_1 \mapsto m_1$ $s_2 \mapsto m_2$		

Table 3.1: Pure admissible strategies in Parikh's (2000) game.

and role of three factors needs to be clarified and their interdependence understood: (i) the true context (nature), which determines the true distribution over states, (ii) individuals' subjective priors, and (iii) the beliefs that interlocutors have about each other's priors.

We begin by sketching out Parikh's analysis of the strategic situation presented in Figure 3.1. This analysis is conducted under static equilibrium methodology (see §2.2) and consists of two main ingredients. First, it is assumed to be common knowledge how likely s_1 and s_2 are; or at least whether one is more likely than the other. Call the probability of the former state p and that of the latter p'. Second, in acknowledgment of the problem that the existence of multiple equilibria poses to static analysis (see §2.3), Parikh proposes a two-step solution concept. First, interlocutors are assumed to identify the game's Nash equilibria, as standard equilibrium analysis would have it. Second, out of the equilibria identified in this manner, interlocutors then pick out the unique remaining Pareto-dominating equilibrium. In informal terms, one outcome Pareto-dominates another if strategies can be changed such that at least one player is better off in the first outcome than in the second without making anyone else worse off.

Let us illustrate these ideas. If we restrict our attention to pure admissible strategies, there are four sender strategies and two receiver strategies in this game. They are listed in Table 3.1. Sender σ_1 sends ambiguous m_3 in s_2; σ_2 sends m_3 in s_1; σ_3 always signals ambiguously; and σ_4 always does so unambiguously. Receiver strategies only differ with respect to how m_3 is interpreted, with ρ_1 taking it to signal s_2 and ρ_2 interpreting it as s_1 instead.

If both state probabilities, p and p', are positive and the sender's preference ordering over messages is as described above, then this game has only two Nash equilibria: $\langle \sigma_1, \rho_1 \rangle$ and $\langle \sigma_2, \rho_2 \rangle$. The first succeeds in using ambiguous but preferred m_3 to signal s_2. The second does so by associating m_3 with s_1. For the receiver either equilibrium is equally good. However, if $p > 0.5$ then the second equilibrium is better for the sender because the least costly message is associated with the most probable state. Under Parikh's assumption of common knowledge about the relation in which p and p' stand, both players are aware (that the other is aware) of this. Consequently, the second-order selection criterion of Pareto-dominance predicts that the optimal outcome $\langle \sigma_2, \rho_2 \rangle$ is played by. The situation is reversed if $p' > 0.5$, with $\langle \sigma_1, \rho_1 \rangle$ being the game's optimal and unique solution. Lastly, in case $p = p'$ the sender is predicted to signal unambiguously by adopting σ_4 because her uncertainty about the receiver's interpretation of m_3 makes ambiguous strategies less attractive than guaranteed information transfer (Parikh 2000:206).[1]

Both ingredients of Parikh's analysis have received criticism. For our purposes, the interpretation of probabilities p and p' is central, but see van Rooy 2004a for detailed criticism of Pareto-dominance as a second-order selection criterion.

From the above, it should be clear that common knowledge of the relation in which p and p' stand is central to ensure coordination. This is the main determinant of the outcome the analysis predicts. What are these probabilities and where do they come from? Parikh (2000:197) notes that they can be "objective or subjective, in general" but that, more often than not, they will be subjective as "objective information will often not be available." On the one hand, Parikh hereby concedes that subjective probabilities matter. Rather than s_1 being more frequent than s_2 or vice-versa, what matters is whether s_1 is (believed by the receiver to be) more likely to be intended than s_2. On the other hand, he suggests that "in the absence of any special information" interlocutors can adduce that if s_1 is more frequent than s_2 (and this is common knowledge), it is also more likely to be intended. How this claim is to be reconciled with the claim that objective information is often not available, as well as what counts as a "special information" is not explained. In light of the centrality of these probabilities and their direct involvement in agents' deliberation processes more should be said about them.

Parikh's analysis ultimately necessitates a frequentist interpretation of p and p'. Otherwise, if these probabilities were purely subjective, the speaker would not experience the difference in payoffs between the two Nash equilibria necessary to employ his second-order selection criterion (Franke 2013:277). More precisely, they may be subjective but this information needs to accord with the true frequency of the states. As argued by Allott (2006:§4), it is therefore important to clarify what the nature of these frequencies is. If, for instance, *frequency of a state* is interpreted as the frequency to which a state is true, then Parikh's analysis seems to make wrong predictions. Intuitively, if frequency is truth based, the utterance *Bolivar wrote a letter* should then be interpreted as (5a) and not as (5b).

(5) a. Bolivar wrote a letter of the alphabet.

[1] Additional ideas that Parikh introduces are needed to fully explain this prediction. After all, with $p = p'$ we face the problem of multiple equilibria again. These ideas are not relevant to what follows but the gist is that the sender then makes her choice relative to the utility she can expect if the receiver is equally likely to pick ρ_1 or ρ_2. That is, she maximizes .5 $\mathrm{EU}_\sigma(\cdot, \rho_1)$ + .5 $\mathrm{EU}_\sigma(\cdot, \rho_2)$, with all sender strategies being on the table again. In this case the best sender strategy is σ_4.

b. Bolivar wrote a letter of correspondence.

I agree with Allott (2006) that these probabilities cannot correspond to truth-based frequency, with Franke (2013) that *something* needs to be interpreted in frequentist terms if it is to have an impact on players' payoffs, and with Parikh himself that, for linguistic behavior at any given point in time, it is subjective probabilities – what interlocutors believe – that ultimately matter. The notions we introduced in Chapter 2 give us the means to coherently put together these requirements.

Technically, the true distribution over states or nature, $P^* \in \Delta(S)$, is what determines state frequencies. Under my preferred interpretation, what determines P^* is the context of interaction. An admittedly artificial but illustrative example is an experimental setup where this distribution is controlled by the experimenter (see §3.5 for concrete illustration). For example, if the task is to describe different objects, the experimenter plays the role of nature in deciding what object a subject is to describe. Viewing P^* in these terms does not imply it being truth-based. Depending on the stimuli of the experiment, (5a) may be more likely than (5b), and vice-versa. Beyond experiments, the information state a calligrapher wishes to convey may skew P^* differently than if she were a secretary; just as a zoologist may be more likely to speak about a *bat* in its zoological sense than a baseball player.

We should stress that, by contrast to the true distribution over states, a prior, $pr \in \Delta(S)$, represents an individual's subjective expectations in a context (see §2.4). The true distribution over states and individuals' subjective priors are different objects. The former determines the state to be conveyed in an interaction whereas the latter determine expectations over states in any such interaction. We had already made this distinction in Chapter 2 but now it should become clearer why keeping P^* and pr apart is important.

The intuition that *Bolivar wrote a letter* is more likely to be interpreted as (5b) implies that this is due to our private contextual expectations in the underspecified context presented above. In a different context we may entertain different expectations. Crucially, and contra Parikh, I do not see why there should be a guarantee that in any context our expectations fully align with P^*; nor with each other's subjective expectations. This is an empirical question. As suggested by the data we analyze in Section 3.5 neither assumption seems warranted.

This brings us back to the questions we set out to address. As noted earlier, many approaches to ambiguity make use of contextual information but it is usually assumed that this information is shared. So far, we discussed Parikh's (2000) analysis. Based on it, we argued that it is important to tease apart the context of interaction from the subjective expectations individual's entertain in it. In what follows, our goal is to elucidate the conditions under which ambiguous signaling can be successful in the face of varying contextual information among individuals, as well as to understand the relation in which such subjective information stands to the true distribution over states.

3.3 Ambiguous Signaling Through Pragmatic Inference

We model language use to the effect that a speaker decides whether to send an ambiguous message based on her beliefs about her interlocutor's likely interpretation of it. That is, speakers gauge whether their addressees will be able to infer the intended meaning from an ambiguous message. If not, they may opt for a less ambiguous one to minimize the risk of misunderstanding. In turn, a hearer's interpretation of an ambiguous message will depend on her subjective expectations in a given context: the relative saliency of interpretations that are truth-conditionally compatible with the message used by the speaker (as discussed in §2.4.2). As before, individuals' expectations in a given context are represented by a prior over states $pr \in \Delta(S)$, where pr^i is player i's prior. Importantly, players are uncertain about their interlocutor's prior. This uncertainty is captured by a distribution over priors, $\mathscr{P} \in \Delta(pr)$. Contrary to past approaches, this information is not common. Put differently, $\mathscr{P}(pr)$ represents a player's belief about pr being her interlocutor's prior.

We propose a conservative generalization of the rational language use model introduced in Section 2.4 to incorporate these assumptions. Player i's literal receiver and sender behavior remain unchanged. They are repeated below as (3.1) and (3.2).

$$\rho^0(s|m; pr^i) \propto L_{[s,m]}\, pr^i(s); \tag{3.1}$$

$$\sigma^0(m|s) \propto L_{[s,m]} - c_\sigma(m). \tag{3.2}$$

Letting L and the sender's preferences over messages be as above, our initial question about the motivations for deliberate ambiguous signaling can be recast as asking under which conditions the risk incurred by the use of preferred m_3 undercuts the benefit of safe unambiguous communication using only m_1 and m_2.

The tension of a sender wanting to uphold her message preferences as much as possible while taking the, possibly diverging, expectations of her interlocutor into consideration arises when higher reasoning types of level $n+1$ are considered. As discussed in Section 2.4, we restrict our attention to boundedly rational agents of level 1. This minimal degree of mutual reasoning suffices to associate m_3 with a salient state under suitable conditions: when the receiver's prior, respectively, the sender's beliefs about it, are informative enough.

Our departure from previous models of rational language use concerns the behavior of the sender, who, instead of using her own prior to anticipate the receiver's behavior, employs her beliefs about the receiver's prior $\mathscr{P}$. Letting θ

codify the parameters of pr, the level-1 behavior of player i is then given by:[2]

$$\rho^1(s|m;pr^i) \propto \exp(\lambda \frac{\sigma^0(m|s)pr^i(s)}{\sum_{s'}\sigma^0(m|s')pr^i(s')}); \tag{3.3}$$

$$\sigma^1(m|s;\mathscr{P}) \propto \exp(\lambda((\int \mathscr{P}(\theta)\rho^0(s|m;\theta)d\theta) - c_\sigma(m))). \tag{3.4}$$

The behavior of speakers of level 1 defined in (3.4) corresponds to the quantal best response to a belief-weighted level-0 hearer. The latter is derived from the domain of $\mathscr{P}$, a set of possible receiver priors, with weights according to the sender's belief in them as corresponding to the actual prior of the receiver.

This proposal is conservative in that it retains the predictions made by previous models of rational language use when the prior is (believed to be) common (cf. Frank and Goodman 2012, Franke and Jäger 2014). This situation is given when $\mathscr{P}$ is degenerate, ruling out all but the speaker's own prior. While in this case a common prior and a belief in a common prior are behaviorally indistinguishable, they are nevertheless conceptually different. The former requires equality of expectations whereas the latter represents beliefs of players about each other's expectations. While a belief in a common prior may often be false, it leads players to behave *as if* there was a common prior. More importantly, this generalization can additionally capture situations in which the sender is either uncertain about her interlocutor's expectations; or is certain but believes that they differ from her own. Such situations can arise in a number of ways but behaviorally boil down to a speaker's increased tendency to use safer messages when uncertain; or to use messages in a way that might go against her own prior but be in line with her beliefs about her addressee's prior, respectively. Reasoning beyond level 1 would allow for further variability in receiver behavior depending on her beliefs about the sender's beliefs, and vice-versa for the sender. We do not make use of such additional layers of complexity here.

To illustrate how (3.4) plays out, let us assume that there are only two distributions in the support of $\mathscr{P}$. For example, $pr_v(s_1) = 0.9 = pr_w(s_2)$. Prior pr_v strongly favors state s_1 over s_2, and vice-versa for pr_w. Furthermore, let interlocutors tend to maximize expected utility (high λ), rendering their behavior more deterministic, and assume that the lexicon and the cost-induced order over messages is as above. While there is a gamut of possible speaker behaviors that arise from an interaction between $\mathscr{P}$ and the concrete values assigned to λ and the cost of messages, there are three general cases of interest. The first is given by $\mathscr{P}$ assigning high probability to pr_v. In this case, ambiguous m_3 is sent in

[2] Alternatively, when considering a finite subset of $\mathscr{P}$'s domain:

$$\sigma^1(m|s;\mathscr{P}) \propto \exp(\lambda((\sum_{pr} \mathscr{P}(pr)\rho^0(s|m;pr)) - c_\sigma(m))).$$

s_1 to maximize expected utility. Since the receiver is believed to expect s_1, m_3 is judged to be risky in s_2. Consequently, unambiguous m_2 is sent in s_2 instead. The second case, in which high probability is assigned to pr_w, is the opposite of the first: m_3 is sent in s_2 but not in s_1, where m_1 is sent instead. Lastly, the sender may be uncertain about the receiver's prior, reflected, for example, by $\mathscr{P}(pr_v) = \mathscr{P}(pr_w)$. In this case, the speaker will opt for the safe strategy of sending m_1 in s_1 and m_2 in s_2.

Single interactions already allow us to quantify how well a pairing of signaling strategies fares in a context. However, the degree of agents' success chiefly depends on their (beliefs about their interlocutors') priors, and on how well these match the context's true distribution over states P^*. A crucial component missing from such an analysis is the possibility of players to interact with each other multiple times. Clearly, if they know nothing about each other, the best a player can do is to make a guess and hope for the best. By contrast, iterated interactions allow senders to change their beliefs according to information obtained from receivers' linguistic behavior, as well as for subjective expectations over states to adapt to the context itself.

Iterated interactions. More often than not communication involves iterated rather than single interactions. This allows interlocutors to adapt to each other. In dialog, linguistic alignment is evinced on many levels: from phonetic (Kim et al. 2011) or syntactic (Pickering and Ferreira 2008) to lexical and referential (Brennan and Clark 1996, Clark and Wilkes-Gibbs 1986, Hawkins et al. 2017). Here, we are concerned with the relation between subjective contextual expectations, beliefs about them, and the information provided by the context in which interactions take place. The latter is codified in P^*, which interlocutors are indirectly exposed to while they interact. What is missing, then, are means for priors and beliefs about them to change over time.

Communication ensues as before. The sender wants to convey a state and sends a message. The receiver interprets it and both players receive a payoff. However, now the players' own subjective priors and their beliefs about their interlocutor's prior are updated based on information gained from the interaction.

There are many ways in which these updates could be modeled. For subjective priors we will assume them to be updated using a simple form of Roth-Erev reinforcement learning (Roth and Erev 1995, Erev and Roth 1998). The motivation behind this rather simple learning mechanism is to (ideally) obtain high rationality outcomes from low rationality behavior (Huttegger et al. 2013). Additionally, it allows us to maintain an analogy to simple biological learning processes (Thorndike 1898, Herrnstein 1970). In human terms this process is akin to priming in that a state's saliency increases as interlocutors are exposed to it (Pickering and Garrod 2004, Reitter and Moore 2014).

More concretely, we assume subjective priors over states to be updated based on a player's *accumulated propensity* for each state s at interaction t, $ap_t(s)$. Accumulated propensity can be likened to a record of the states that the sender intended to communicate, or the receiver interpreted, in previous interactions. The propensity for the state in play is updated by a value r after an interaction. This value is positive in case of communicative success and negative in case of failure, as reflected by $\delta(\cdot, \cdot)$ (see definition (2.3)). For sender i that sent m in s with receiver j interpreting this message as s' this gives:

$$\begin{aligned} ap^i_{t+1}(s) = ap^i_t(s) + f(r), \text{ where } f(r) &= r && \text{if } \delta(s, s') = 1, \\ f(r) &= -r && \text{otherwise.} \end{aligned}$$

The receiver's accumulated propensity for state s' that she took m to signal, $ap^j_{t+1}(s')$, is updated analogously.

Before interacting, player i's propensity is simply proportional to her prior, $ap^i_0(s) \propto pr^i(s)$. Subsequently, player i's prior for interaction $t+1$ is derived from her amassed propensity up to interaction t: $pr^i_{t+1}(s) \propto ap^i_t(s)$.

The value by which propensities change controls how fast the initial prior is overridden. Small r gives the initial prior more weight whereas larger values lead players to abandon or reinforce their preconceptions faster. Negative reinforcement is not required for the results reported below to obtain. However, it speeds up the process. Other possibilities include the addition of recency effects – by weighting recent states higher than less recent ones – or learning with suppression – by decreasing the association strength of states that were not in play (Erev and Roth 1998, Franke and Jäger 2011). Alternatively, interlocutors could use more sophisticated mechanisms to update their priors. As with our previous choices, we decide for a simple and transparent mechanism that serves our purpose. The contribution of reinforcement learning to our following predictions is straightforward and can be achieved in a number of ways: a player's expectations of a state should grow with increased exposure to it.

Note that r is dissociated from utility. This diverges from most previous signaling models with adaptive learning dynamics (e.g., Barrett and Zollman 2009, Franke 2016). Notwithstanding, this assumption is warranted here as there is no reason to relate a speaker's prior over states to incurred production cost. In fact, a direct association between utility and prior updates would have undesirable consequences in cases where messages true of less frequent states are less costly than those true of more frequent ones. This could lead to the former being more salient than the latter. In informal terms, having a preference to talk about something in a particular fashion should not make it a priori more probable to be spoken about.

In contrast to the somewhat mechanistic fashion in which subjective priors are updated, we assume the change of a sender's beliefs about her addressee's prior to involve an inferential component. Here, we model it as an update of $\mathscr{P}$ that consolidates old with new information using Bayes' rule. This reflects the sender's primary goal to actively reach understanding by correctly anticipating her addressee's interpretation. This motivation can already be seen as rooted in agents' engagement in mutual reasoning.

The evidence witnessed by the sender on which she bases her inference is whether communication succeeded. However, she receives no information about the receiver's interpretation if communication failed, beyond the fact that it failed.[3] More precisely, in an interaction in which the speaker wanted to convey s with message m, interpreted as s' by the receiver, the sender witnesses $w(s)$, where $w(s) = \{s\}$ if $\delta(s, s') = 1$. Otherwise $w(s) = S \setminus \{s\}$. Based on evidence $w(s)$, the sender adjusts her beliefs about her interlocutor's prior based on the likelihood of a prior leading to the witnessed receiver behavior. Accordingly, $\mathscr{P}$ is updated as follows:

$$\mathscr{P}_{t+1}(pr \mid w(s); m) \propto (\sum_{s' \in w(s)} \rho^0(s' \mid m; pr))\mathscr{P}_t(pr). \tag{3.5}$$

When interacting again, linguistic choice is computed as before with updated priors and updated beliefs over them.

3.4 Predictions for Single and Iterated Interactions

Based on the preceding discussion, a straightforward first prediction is that ambiguous communication is at least functionally equivalent, in terms of information transfer and fulfillment of speaker preferences, to unambiguous counterparts provided that (i) the speaker's beliefs about the receiver's prior correctly anticipate her actual behavior, and that (ii) signaling behavior is relatively deterministic. Condition (ii) is important to ensure that receivers have a tendency to associate ambiguous messages with a single state in a given context. More importantly, under these conditions ambiguity is functionally advantageous when there are at least two contexts governed by distributions over states that each assign a non-zero probability to distinct states associated with a preferred ambiguous message, and the speaker uses this message in both contexts. The existence of multiple contexts is important because for every single context there is an unambiguous lexicon that fares at least as well as an ambiguous one. For example, one in which m_3 is only true of s_1 if $P^*(s_1) \geq P^*(s_2)$ or one in which this message is true only of s_2 if $P^*(s_1) \leq P^*(s_2)$. An advantage for ambiguity can therefore only manifest

[3] This is one of the main differences between signaling games in the Lewisian tradition and so-called *naming games* (e.g., Steels 1998). In the latter tradition it is standardly assumed that the true state of the world (or intended referent) is revealed to the receiver after each interaction. For a game with two states this makes no difference: even when players fail to communicate there is only one other state the sender could have intended to convey.

when there are multiple contexts. We return to this matter in Section 3.6 and address it in detail in Chapter 5. Lastly, ambiguity is maximally advantageous in a context if the most frequent state in it is associated with the least costly message. Put differently, the most frequent state(s) in a context ought ideally be associated with the most preferred form(s) when speaker economy is at stake.

The adoption of an ambiguous strategy ultimately hinges on the sender's beliefs about the receiver. Whether the aforementioned advantages manifest therefore depends on factors that would lead agents to have similar expectations (over expectations). This also means that ambiguous signaling is more risky in a world in which contextual expectations greatly vary across agents. In the case of humans, behavioral experiments suggest that they generally succeed, at least significantly beyond chance, in matching their expectations with those of others when it is known that the other party is trying to do the same (Schelling 1960, Mehta et al. 1994; see Section 3.5 for data particular to ambiguous signaling as well as the discussion on focal points in Section 2.4.1). However, from previous accounts and our analysis so far, it is unclear how agents may come to entertain such aligned expectations.

That speakers are somewhat aware of the contextual expectations of their interlocutors is also evident in the use of puns, such as (6) - (8), which are funny because they manipulate and exploit the likely expectation of how ambiguity will be resolved.

(6) What is the difference between a hippo and a Zippo? One is very heavy and the other is a little lighter.

(7) Why did the man fall in the well? Because he could not see that well.

(8) Two goldfish are in a tank. One turns to the other and says, "You man the guns, I will drive".

To recapitulate, ambiguity can be advantageous in single interactions as long as sender beliefs anticipate receiver behavior. Crucially, subjective priors need not match for ambiguity to be exploited. No common prior is required. On a general level, this characterization is nevertheless in the spirit of previous justifications of ambiguity. The explanatory burden shifts from a common prior to sufficiently accurate beliefs about the receiver's prior. However, this shift highlights that the conditions for safe ambiguity exploitation may not always be given and allows us to ask when and how they can be reached. Whether an ambiguous signal is understood depends on the receiver's own expectations; whether it is sent depends on the sender's beliefs about these expectations; and the utility of conveying a particular state by an ambiguous message will – in the long run – depend on the true distribution over states. We now turn to iterated communication to tease apart the interaction between these factors and to elucidate how and under which conditions an advantage crystallizes.

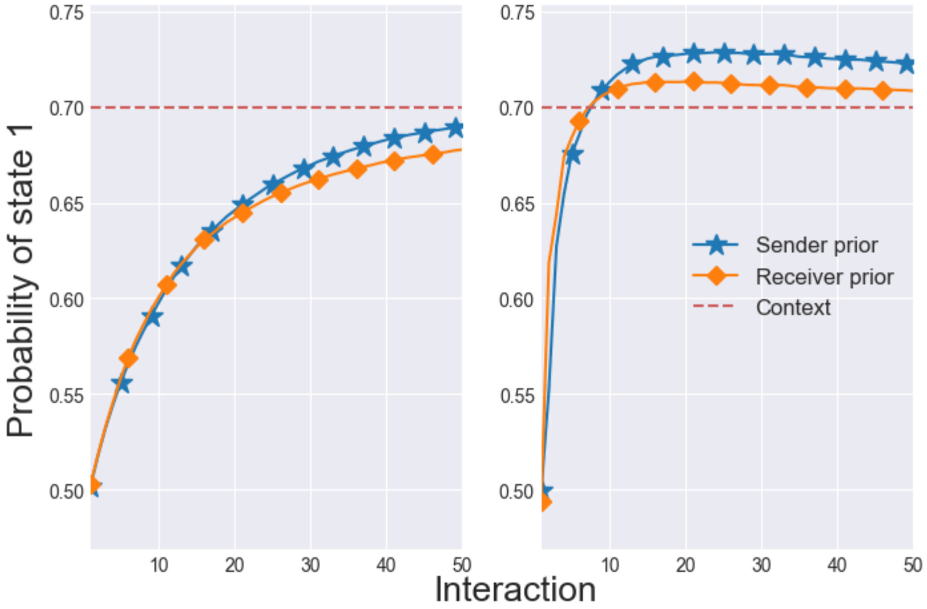


Figure 3.2: Mean subjective prior development in 10^4 independent simulations with $P^*(s_1) = 0.7$ for $r = 0.1$ (left) and $r = 1$ (right).

3.4.1 Simulations

In order to inspect the model's predictions in more detail, a sender's initial beliefs about the receiver's prior need to be set. Here, we assume sender i's initial $\mathscr{P}$ to be Dirichlet distributed, with weights for state s set to $q \times pr^i(s) + 1$. In words, high q corresponds to the sender believing that the receiver's expectations are close to her own, with $q \to \infty$ approaching the belief of a common prior. Lower values correspond to more divergence and uncertainty. In the extreme, $q = 0$ corresponds to complete uncertainty about the receiver's prior; every prior is deemed equally probable.

For the following simulations, we use L as above and assume that $\lambda = 20$, $c_\sigma(m_1) = 0.4 = c_\sigma(m_2)$ and $c_\sigma(m_3) = 0.1$. That is, players are subjectively rational but might occasionally fail to maximize utility (from their subjective perspective), and ambiguous m_3 is preferred over either unambiguous message, each of equal cost. To inspect the average outcome of interactions, including best- and worse-case scenarios, players' priors are randomly sampled at the onset of a first interaction. The sender's value for q is randomly sampled from $[0; 20]$ at the onset as well.

The mean development of players' subjective prior when interacting in a context governed by $P^*(s_1) = 0.7$ is illustrated in Figure 3.2 for two values of reinforcement parameter r. This figure shows that priors approach the true distribution of the context as players interact in it. When there are only two states this simple learning process is particularly fast because negative reinforcement in one

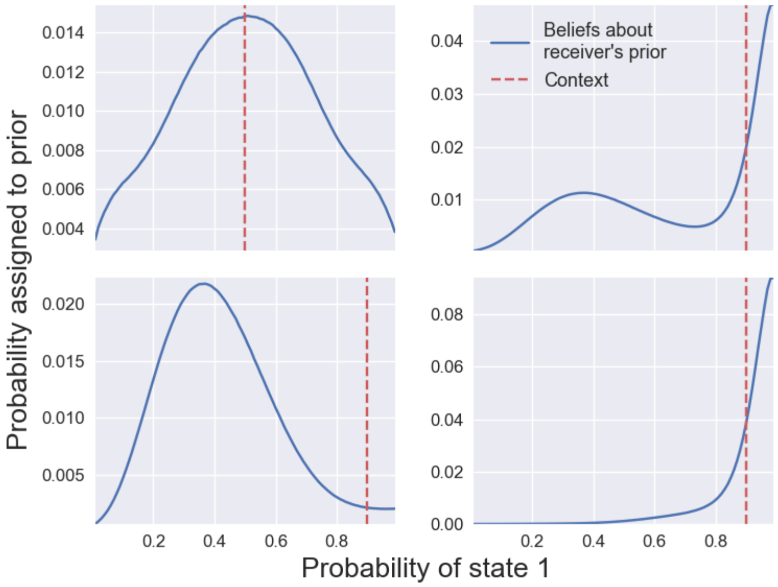


Figure 3.3: Mean beliefs about receiver prior in 10^4 independent simulations with $P^*(s_1) = 0.5$ (top-left) and $P^*(s_1) = 0.9$ (top-right). The lower row splits the mean beliefs of senders in $P^*(s_1) = 0.9$ by their expected sender utility being less or equal than 0.63 (bottom-left) and greater than 0.63 (bottom-right). The latter are beliefs of senders that, at least in tendency, associate s_1 with m_3. The former do not.

state leads to the prominence of the other. Figure 3.2 also showcases the role of r in controlling the speed by which priors converge to a context's distribution over states.

In the following we focus on results obtained from an r-value of 0.5 after 50 interactions. The latter ensures that the reported outcomes approximate endpoints of the dynamics but should not be taken as indicative of the minimal number of interactions required to reach them. Supplementary results obtained from less interactions and different r-values are provided in Appendix A.

A more central interaction is that between P^* and a sender's beliefs about her interlocutor's prior, as well as their bearing on the choice of ambiguous m_3. Figure 3.3 showcases how the context influences sender beliefs. The top row shows the beliefs of senders about their interlocutors' priors after 50 interactions for two true distributions over states. The bottom row zooms in on the sender's beliefs in $P^*(s_1) = 0.9$; they show the difference in the beliefs of senders that (at least in tendency) adopt the optimal strategy of associating m_3 with the most frequent state (right) and that of those that opt for safe unambiguous signaling instead (left). Different contexts give rise to similar patterns, with a belief-peak centered around 0.5 in the case of unambiguous signalers and peaks that tend to the extremes of the space in the case of ambiguous ones.

Recall that $\mathscr{P}$ is updated based on what can be inferred about the receiver's prior from her behavior. The only interactions that are informative about this matter and therefore influence a sender's beliefs are the receiver's interpretations of ambiguous messages. In turn, the receiver's interpretation of an ambiguous message may change over time due to her exposure to the context (see Figure 3.2). In particular, contexts that are not very informative can lead to fluctuations in the receiver's expectations, making her interpretative behavior more difficult to predict for the sender. Consequently, as showcased by the top-left plot in Figure 3.3, senders grow uncertain about their interlocutor's expectations in such contexts. The uninformative prior that receivers converge to in such contexts does not lend itself for the safe exploitation of ambiguity either. Uncertainty about expectations centered around uninformative priors therefore often lead to the avoidance of risky signals. By contrast, receiver expectations in contexts in which one state is markedly more frequent than the others are fairly predictable after a few interactions. Senders pick up on this fact once they employ an ambiguous signal.

As shown for $P^*(s_1) = 0.9$ in the right plots of Figure 3.3, senders tend to overestimate their interlocutor's prior in contexts that strongly favor one state. This is due to the likelihood of a correct interpretation of m_3 being higher the more degenerate subjective priors are. Overestimation decreases as mutual reasoning levels increase but predictions about the use or avoidance of ambiguous messages do not hinge on the shape of the sender's belief but on the range of priors it concentrates on. That is, a false belief about an addressee's prior is not detrimental to communication if it correctly predicts behavior.

The amount of senders that adopt an ambiguous strategy in a context is reflected most clearly by their expected utility (see definitions (2.4), (2.5) and (2.8)). An excerpt of the mean expected sender utility together with the mean Jensen-Shannon divergence (JSD) between the interlocutors' priors is given in Table 3.2. Informally, JSD measures the closeness of two distributions as a divergence to their average. More precisely,

$$\text{JSD}(pr_v, pr_w) = {}^1\!/\!_2\, D(pr_v \;||\; M) \;+\; {}^1\!/\!_2\, D(pr_w \;||\; M), \tag{3.6}$$

where

$$M = {}^1\!/\!_2(pr_v + pr_w); \tag{3.7}$$

$$D(pr_v \;||\; pr_w) = \sum_s pr_v(s)\ \log \frac{pr_v(s)}{pr_w(s)}. \tag{3.8}$$

As shown in Table 3.2, even in a context governed by a uniform distribution the mean expected utility of senders is higher than 0.6. This is the value guaranteed by the use of only unambiguous messages m_1 and m_2, irrespective of P^*, for $c_\sigma(m_1) = 0.4 = c_\sigma(m_2)$. The senders' mean expected utility is also always higher

$P^*(s_1)$	EU_σ (SD)	JSD	$\text{EU}_\sigma^{\max}$
0.5	0.61 (0.06)	0.004	0.75
0.7	0.68 (0.10)	0.002	0.81
0.9	0.72 (0.13)	0.002	0.87

Table 3.2: Mean sender expected utility and JSD of interlocutors' priors after 50 interactions in 10^4 independent games per P^*. $\text{EU}^{\max}$ indicates the maximum expected utility reachable for a given P^*. SD is the standard deviation.

than the mean utility of approximately 0.57 expected in the first interaction. The latter value is lower than the safe guaranteed value of 0.6 because priors and q were sampled randomly. This inevitably led to the failure of some initial attempts to exploit ambiguous m_3. As suggested by Figure 3.2, iterated interactions also strongly improve upon the mean initial JSD of approximately 0.15. Lastly, the increase in standard deviation of expected utility with the frequency of s_1 is a consequence of the ensuing increasing difference between the expected utility of an ambiguous signaling strategy against that of adopting an unambiguous one.

Finally, Figure 3.4 shows the proportion of communicative failures across interactions for a sample of true distributions over states. As is to be expected, the initial error rate effected by sampling priors and beliefs about them decreases as the number of interactions increases. Once again, coordination is aided by increase in frequency of one state over others. Contexts with a markedly more frequent state lead to more informative priors. More informative priors are less prone to change and effect more deterministic linguistic behavior.

In sum, this model (i) generalizes past analyses of ambiguity by relaxing the assumption of a common prior, (ii) shows how agents may come to entertain (beliefs about) contextual expectations that allow for the safe exploitation of ambiguity, (iii) highlights the role of context frequencies in enabling or preventing such exploitation, and (iv) connects this research with claims in the alignment literature about its role in dialog optimization, providing interlocutors with means to establish patterns of language use better tailored to the context of their interaction and their preferences (Clark and Wilkes-Gibbs 1986, Reitter and Moore 2014).

More broadly, this interactive perspective also highlights the function of ambiguity as an opportunistic adaptive device that endows agents with the ability to mold language use to their interlocutors and the environment, and links this opportunism to the information provided by a context. Contexts of high informativity are particularly conducive to ambiguity exploitation because they (i) foment less fluctuations in the receiver's interpretation of ambiguous messages and, consequently, (ii) lead to less uncertainty in the sender's beliefs about her interlocutor's prior. The expected utility of senders also increases with context informativity as these contexts more often lead to the association of frequent

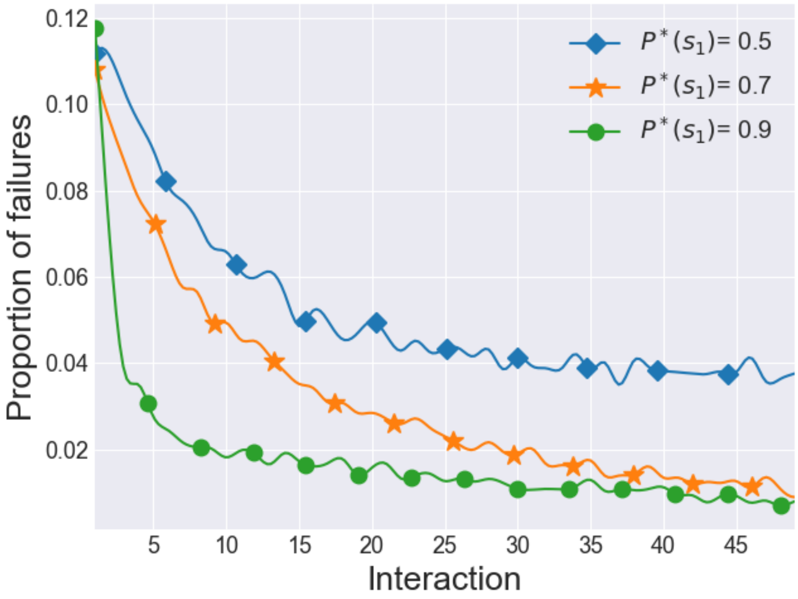


Figure 3.4: Proportion of communicative failures across 50 interactions in 10^4 independent simulations per true distribution over states.

states with preferred but ambiguous messages. By contrast to, e.g., Parikh 2000, this association is not explicitly sought after by senders but rather is a byproduct of a receiver's association of an ambiguous message with its most salient interpretation. The interplay of saliency, frequency and interpretation therefore often leads players to adopt Pareto-optimal signaling strategies.

Nevertheless, as indicated by Table 3.2, not even informative contexts guarantee that the optimal ambiguous strategy is always adopted. There are two intertwined reasons for this. First, we allowed the priors of interlocutors to vary freely before engaging in communication. This may cause a speaker with an uninformative prior and high q to believe her interlocutor's prior to be uninformative as well. Consequently, such a speaker will never try to use an ambiguous message even after exploring the context (which may turn out to be informative). Similarly, initial uncertainty from low q may lead speakers to not use risky signals, meaning that they never learn anything about the receiver's expectations. Second, a great number of interactions started with opposing contextual expectations. This can lead to an early communicative failure when using an ambiguous message. As in the other cases mentioned, this can then deter the sender from using risky messages in the future. As we will see once we look at the empirical data in Section 3.5, the fact that the model does not always converge to the game's optimal outcome but instead allows for unambiguous strategies or Pareto-dominated ones to entrench themselves is a desirable feature. Nevertheless, we briefly explore two ideas that qualify whether these situations pose challenges to successful coordination with ambiguous messages.

3.4.2 Exploration and past experience

Communication draws from past experience and agents may often find themselves in similar contexts. This enables visitors of zoos and baseball courts alike to use plain *bat* without first probing whether their interlocutor is attentive to the same meaning. They have experience in these contexts and assume that their interlocutors have had some too; at least to a degree to which one interpretation of ambiguous *bat* is markedly more expected than the other. Once we allow for richer background knowledge of a context the issue of strongly diverging initial priors is reduced. A shared cultural background and experience in an environment may therefore suggest themselves as partial answers to the question how linguistic coordination with ambiguous messages can succeed prior to multiple interactions.[4]

The question how the speaker's initial q-value is determined remains, however. While a detailed treatment is outside the scope of this chapter, one possibility is for it to be sensitive to the informativity of a context in combination with beliefs about the receiver's experience in similar ones. In broad strokes: high q may come about because the context is assumed to be well known. Either because this is known about the receiver itself or because this context is common enough that members of a population are taken to be familiar with it. An informative context that is assumed to have been encountered frequently enough may then lead to an optimistic speaker strategy in which ambiguity is believed to be (usually) resolvable (cf. Clark and Schober's (1992) *presumption of interpretability*). Even for such optimistic speakers, adaptive dynamics would still play a role in unknown or infrequent contexts, as well as as corrective devices when optimism turns out to be misplaced.

3.4.3 Preemptive adaptation

Next, we turn to the issue of senders who, due to early communicative failure or initial uncertainty about their interlocutors' expectations, remain averse to ambiguity even in informative contexts. The reason for senders occasionally locking-in on an unambiguous strategy even if they could safely exploit ambiguity is that the update of $\mathscr{P}$ is not sensitive to the information gained from the context; nor to the fact that interlocutors adapt to it over time. There are different alternatives that can stimulate the exploration of ambiguous strategies after learning more about the context. A simple one is for $\mathscr{P}(pr)$ to be affected by the probability of the current state s under pr. To this end, we can keep our update mainly as

[4] There are many ways in which this idea could be implemented. For example, initial priors could be derived from samples from P^*, or from past interactions with other agents. We chose not to do so as we hope the positive effect this idea would have are clear.

$P^*(s_1)$	EU_σ (SD)	JSD	$\text{EU}_\sigma^{\max}$
0.5	0.58 (0.114)	0.033	0.75
0.7	0.8 (0.022)	0.001	0.81
0.9	0.87 (0)	0	0.87

Table 3.3: Mean sender expected utility and JSD of interlocutors' priors after 50 interactions in 10^4 independent games using "preemptive" belief updates. $\text{EU}^{\max}$ indicates the maximum expected utility reachable for a given P^*.

it was in (3.5) and simply add $pr(s)$ as a last term:

$$\mathscr{P}_{t+1}(pr \mid w(s); m; s) \; \propto \Big(\sum_{s' \in w(s)} \rho^{n-1}(s' \mid m; pr)\Big)\mathscr{P}_t(pr)\, pr(s). \tag{3.9}$$

This operationalizes a sender that changes her beliefs on the assumption that her interlocutor adapts to the context; "preemptively" exploiting the relative saliency of states that were relevant before. Table 3.3 shows how this modification affects the outcome of interactions. As with the update in (3.5), supplementary results obtained from less interactions and different r-values are provided in Appendix A.

In a nutshell, this less conservative update fares well in contexts governed by distributions that favor a single state but less in those governed by flatter ones. In the former kind of context the proportion of dyads that adopt the Pareto-optimal strategy of associating m_3 with the most frequent state is markedly higher than under the simpler update mechanism in (3.5). However, as shown for $P^*(s_1) = 0.5$, this can come at a cost in less informative contexts. The modified update in (3.9) favors priors that are informative about the current state in play. Consequently, while the receiver converges to a prior that is not well-suited for ambiguity exploitation in uninformative contexts, speakers instead tend to infer more informative priors, attempt the use of risky signals, and often fail. By contrast, our main proposal for updating $\mathscr{P}$ leads to more cautious behavior that may not always result in ambiguity exploitation but generally ensures that communication succeeds.

3.5 Model Fit

In a recent study, Kanwal et al. (2017) use an artificial language learning paradigm to investigate people's tendency to associate short forms with frequent meanings. The experiment's *combined condition*, detailed below, closely resembles the situation we analyzed above: there is a preference for an ambiguous message over either of two unambiguous alternatives; the communicative task subjects are involved in is cooperative; and there is a latent context that controls state frequencies, with one object being more frequent than the other. Crucially, subjects had no direct access to the latter information. Instead, they were implicitly exposed to it throughout the experiment. In our notation, this context is such that $P^*(s_1) = {}^{24}\!/_{32} = (1 - P^*(s_2))$, with s_1 denoting the experiment's frequent object and s_2 the infrequent one. After surveying Kanwal et al.'s setup, we use their data to assess how well our model can explain subjects' behavior and use our fit to explore the data.[5]

In contrast to other studies that use iterated coordination tasks in the form of dialog (e.g., Krauss and Weinheimer 1964, Clark and Wilkes-Gibbs 1986), the vocabulary available to subjects in this study was not open ended. Instead, participants were trained on an "alien" language, consisting only of three names and two objects. One name, *zop*, was ambiguous and could apply to either object. The other two names, *zopudon* and *zopekil*, were unambiguous and applied only to one object each. Restricting language use to only these three forms ensured that the two objects stood in direct competition for the single short label *zop*. The artificial language that subjects were trained to use to communicate with each other was accordingly one of the following two:

$$\begin{array}{c} \\ object_1 \\ object_2 \end{array} \begin{array}{c} \begin{array}{ccc} zopudon & zopekil & zop \end{array} \\ \left[\begin{array}{ccc} 1 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right] \end{array} \qquad \begin{array}{c} \\ object_1 \\ object_2 \end{array} \begin{array}{c} \begin{array}{ccc} zopudon & zopekil & zop \end{array} \\ \left[\begin{array}{ccc} 0 & 1 & 1 \\ 1 & 0 & 1 \end{array}\right] \end{array}$$

The experiment consisted of a training phase and a testing phase. In the training phase subjects were iteratively presented with object-label pairs that were true in the artificial language assigned to them. In this way the subjects learned the language that was later used for actual communication with other participants in the testing phase. Training consisted of 32 learning trials, with one object appearing 24 times and the other one appearing only 8 times. After completing the training phase, subjects were assigned to one of the four experimental conditions of the testing phase.

In the *combined condition* subjects were paired to play a signaling game with partners trained on the same language as themselves. Partners were also exposed to the same object frequencies. In each testing trial the sender was presented with one of the two objects, and with a choice to send one of the two labels that applied to this object. For example, if trained on the left language from above, a sender trial could consist of communicating *object₁* with the choice to send either unambiguous *zopudon* or ambiguous *zop*.

In this condition sending messages cost time. To signal, senders had to keep a transmission box pressed while the letters of the selected message appeared one by one. A message could only be transmitted once all letters had appeared. Sending ambiguous *zop* therefore took less time than sending an unambiguous message.

[5] I am extremely grateful to Jasmeen Kanwal for making this data available to me and answering my inquiries about it. This analysis would not have been possible without her help.

The receiver's task, upon receiving the message, was just to select the intended object from an array showing both *object$_1$* and *object$_2$*. After receiving feedback about their communicative success the subjects' roles reversed. In this way, each participant played a total of 32 sender and 32 receiver trials with their assigned partner. Similarly to the training phase, each sender had to communicate one object 24 times and the other one only 8 times. The frequencies of the objects in the testing phase matched those that subjects had already been implicitly exposed to in their training. That is, if *object$_1$* was frequent in training then it was also the frequent object in testing. At the beginning of the game subjects were told that the quickest dyad with the highest amount of correct trials would get a reward. This was done to incentivize communicative success as well as a preference for the faster ambiguous message *zop*.

The other three experimental conditions differed from the combined condition in that the pressure for communicative success, the time difference between messages, or both of these factors were removed. One condition had instantaneous message transmission. Another had subjects play a labeling task alone in order to remove the communicative element of the game while keeping the difference in transmission time between messages. The remaining condition had them play alone and with instantaneous messages.

Overall, Kanwal et al. (2017) found that subjects showed a tendency to use unambiguous labels when pressured for communicative success in the condition with no difference in transmission time between short and long messages; to use the short ambiguous label when a time differential existed but no communicative element; to use unambiguous labels when neither time differences nor communication existed;[6] and to use the (un)ambiguous message to communicate the (in)frequent object when there was a time difference and a communicative element. While these tendencies are robust, there was also substantial variation across subjects and trials. Of particular interest to us is that, in the combined condition, some subjects deviated from the optimal strategy of associating the frequent object with the short ambiguous form. Instead, they signaled either unambiguously or associated the infrequent object with the short form.

3.5.1 Individual-level data and model

Kanwal et al.'s (2017) data set for the combined condition comprises the linguistic behavior of 40 participants (20 dyads). As we focused on level-1 senders rather than more complex interactions between higher order sender/receiver beliefs, our

[6] That is, even with a pressure for communicative success removed, subjects showed a tendency to send unambiguous messages when they were as fast to transmit as ambiguous ones. This result is noteworthy in that it suggests that, all else being equal, unambiguous messages are preferred over ambiguous ones. While the results of the combined condition of Kanwal et al. 2017 adds to the evidence against ambiguity aversion in language use, this condition indicates that a baseline preference for its avoidance might exist.

analysis will focus on the 32 sender trials of each subject. We begin by surveying the data to give a broad first impression of how subjects behaved.

Data

Out of a total of 1280 sender trails, 1201 trials were successful (93.8%). Out of the 40 senders, 19 senders managed to communicate the objects successfully in all of their 32 trials. Six failed once; four failed twice; three failed three times; two failed four times; one failed five times; one failed seven times; two failed eight times; and two failed ten times. The trials of five subjects therefore account for more than half of the communicative failures in this experimental condition. Put differently, most subjects did well but, out of those that did not, some accrued a large amount of communicative failures.

To get an idea of how individuals behaved in the experiment, we categorize subjects according to whether they predominantly (i) associated the frequent object with the short ambiguous message, the *Horn strategy*, (ii) associated the infrequent object with the short message, the *Anti-Horn strategy*, (iii) avoided ambiguity, the *unambiguous strategy*, (iv) employed none of the preceding strategies, the *variational strategy*, or (v) failed to communicate more than three times, the *erratic strategy*.

The first two strategies owe their name to Horn's (1984) division of pragmatic labor, which states that (un)marked expressions typically are associated with an (un)marked interpretation. For instance, utterances (9) and (10) are classically analyzed as having the same truth-conditions. However, in virtue of (9) being less marked, it is taken to pragmatically convey an unmarked interpretation. Reversely, uttering (10) signals that there is a reason why (9) was not uttered. This gives rise to a marked interpretation.

(9) Mercader killed Trotsky.
$\rightsquigarrow$ Mercader killed Trotsky in a stereotypical way.

(10) Mercader caused Trotsky to die.
$\rightsquigarrow$ Mercader killed Trotsky in a non-stereotypical way.

In this case, the Horn strategy is the strategy that associates unmarked *zop* with the frequent object. The Anti-Horn strategy uses the reverse signaling pattern, going against Horn's prediction. This is the ambiguous Pareto-dominated strategy that, e.g., Parikh (2000) predicts to never arise.

By contrast to the other four categories, the erratic category gives a crude approximation of the amount of "problematic" subjects (more details below). Given the relative simplicity of the task and the fact that, in principle, safe communication was always an option, we categorize subjects as erratic if they failed to communicate more than a total of three times.

As for subjects that failed to communicate three times or less, we categorize them as (Anti-)Horn if they used ambiguous *zop* to communicate the (in)frequent object in all but at most three occasions. This allows for some deviation from full adherence to these strategies. Similarly, we categorize subjects as unambiguous if they used *zop* at most three times. Senders that did not fulfill any of the aforementioned conditions are categorized as being variational. In total, there were 11 Horn, 4 Anti-Horn, 7 unambiguous, 10 variational, and 8 erratic senders.

Model

Our goal is to see whether our model can explain subjects' individual choices across their sender trials in the testing phase. To do so we need to specify the choice and update mechanisms assumed to underly these choices. Recall that linguistic behavior is driven by the rationality parameter λ, preferences over messages, and a sender's beliefs about her interlocutor's prior, with q and a sender's initial prior determining $\mathscr{P}$'s initialization. We fix preferences over messages with the same values as before. That is, $c_\sigma(\text{zop}) = 0.1$ and $c_\sigma(\text{zopudon}) = 0.4 = c_\sigma(\text{zopekil})$; we use the simple update mechanism defined in (3.5); and also fix the initial prior of all subjects to $pr(\text{frequent object}) = {}^{24}\!/_{32}$. The latter corresponds to the exposure to the true distribution P^* that subjects had after the training phase. As for q, we assume players to come in one of two types: cautious type τ_c or risky type τ_r. We let these types correspond to a q-value of 0 and 40, respectively. This has the effect that cautious type τ_c senders initially believe every receiver prior to be equally probable, making the use of less ambiguous messages more likely due to uncertainty. Risky type τ_r senders instead initially believe their interlocutor's prior to approximate their own prior of $pr(\text{frequent object}) = 0.75$.

In principle, the model would allow us to estimate, for each subject: its own preferences over messages, its initial prior, and its particular q-value. However, our goal is not to show that a model with many parameters can account for subjects' behavior. Instead, we want to see whether we can do so with a constrained and informed model, allowing us to enrich our preceding theoretical analysis by its success as well as its failure. What is more, for our predictions to be interpretable in a meaningful way, the model has to be restricted in some respects. Otherwise, completely unambiguous behavior could for instance be effected either by low q; a reversal of preferences over messages; or an uninformative initial prior. Of course, it is probable that subjects entertained a wider range of beliefs about their interlocutors than captured by types τ_c and τ_r; that they had diverging initial priors even if they were trained in the same way; and that their incentive to use the ambiguous label differed. However, lacking precise information about these matters, we follow the original experimental setup as closely as our model allows while fixing unknowns as best as we can.

In sum and with these provisos in mind, we want to see whether subjects' behavior can be explained by one of two types, τ_c and τ_r together with an individual's degree of soft-maximization λ. This is done under the assumption that subjects are cooperative; that they form beliefs about their interlocutors' contextual expectations based on their linguistic behavior; that they have a preference for ambiguous *zop*; and that they have an initial prior that favors the frequent over the infrequent object. Reversely, failures to explain the data will also shed light on these assumptions.

For each individual i we want to explain 32 data points of the form d^i_{opjkln}. Each such datum codifies a sender's choice in a particular trial as well as the information available to her from past trials to update $\mathscr{P}$. The superscript labels the individual, $i \in [1; 40]$. Index $o \in \{1, 2\}$ codifies whether the subject had to communicate *object₁* or *object₂* in this trial. Index $p \in \{1, 2, 3\}$ codifies the message the subject sent, with 1 and 2 standing for the unambiguous messages true of the object with the same index, and 3 for ambiguous *zop*. Indices $j, k, l, n \in [0; 31]$ are counts of previous successes and failures using the ambiguous message. This information feeds into the sender's beliefs about her interlocutor's prior as specified in the update rule in (3.5). This, in turn, has a bearing on her choice as specified in the choice rule in (3.4). Index j is the count of previous successes communicating *object₁* using the ambiguous message and k that of previous failures to communicate *object₁* using this message. Analogously, l and n are counts of previous successes and failures to communicate *object₂* using the ambiguous message. The likelihood of datum d^i_{opjkln} is then given by the probability P^i_{opjkln} that individual i of type $\tau^i \in \{\tau_c, \tau_r\}$ with λ^i sends message p in state o, conditioned on the counts in $j - n$.

We assume an uninformative prior over types, $\tau^i \sim \text{Bernoulli}(p)$ with $p = 0.5$, and that λ is positive and sufficiently high, $\lambda^i \sim \text{Gamma}(\text{shape} = 30, \beta = 1)$. In words, we expect subjects to be of one of the two types without a bias toward either, and to be sufficiently rational as to exploit the ambiguous message if it is believed to be understood by her interlocutor (initially due to her own type; later due to the information she gets from counts $j - n$).

We expect the model to be able to explain the behavior of subjects that fall into the first three categories reasonably well. That is, that of those that approximate either Horn, Anti-Horn or unambiguous strategies. Furthermore, we expect it to be less successful in explaining the behavior of subjects that experienced multiple communicative failures.

3.5.2 Results

Some of the terminology used in this section presupposes basic familiarity with Bayesian modeling. Readers lacking this background will hopefully nevertheless be able to appreciate the results (see, e.g., Lee and Wagenmakers 2014 for an introduction). Intuitively, what we want to infer is the degree to which a subject maximizes expected utility from her subjective perspective, λ, together with whether she is (initially) cautious or risky. We do so based on the data she produced across her 32 sender trials. We provide the model with a reasonable

	Horn	Anti-Horn	Unambiguous	Variational	Erratic	Population mean
λ	31.34	22.39	25.99	22.066	12.97	23.52
τ_c	0.09	1.0	1.0	0.63	0.31	0.52
τ_r	0.91	0.0	0.0	0.37	0.69	0.48
RMSE	0.13	0.19	0.14	0.24	0.37	0.21

Table 3.4: Mean marginal posteriors and root-mean-square errors.

starting point, the Gamma and Bernoulli distributions mentioned above, and then explore the parameter space to find a good estimate that accords well with the data; inasmuch as the model is able to account for the subject's behavior in the first place. These estimates are not single points but instead yield uncertainty over parameters. We can then use these estimates to predict a subject's behavior, which is a reasonable way to check whether the model can account for the data in the first place. Note that the model knows nothing about our five categories. These are just intended as aid to interpret model and data by providing some coarse-grained bins to classify subjects.

To obtain posterior estimates for our parameters, λ and τ, we fit individuals' data with PyMC3 (Salvatier et al. 2016). For each individual we collected 2000 samples from two chains from the joint posterior distribution after a burn-in of 800 samples. All simulations conducted in this way had an $\hat{R}$-value below 1.05. This suggests chain convergence (Gelman and Rubin 1992).

Mean summary statistics of the marginal posteriors estimated for types and λ-values are given in Table 3.4. Individual estimates for λ with respective highest density intervals are provided in Appendix B.

We also conducted posterior predictive checks by drawing 500 samples from subjects' posteriors for each of their trials. The root-mean-square error (RMSE) between actual values and the ones predicted by the model in this way are also given in Table 3.4 (see Appendix B for individual RMSEs). We showcase a selection of predicted values in more detail below.

Considering the categorical nature of the data and that we have a single data point per trial choice, the model provides a good fit for most subjects. Particularly for those we had categorized as Horn, Anti-Horn and unambiguous. The higher RMSE for Anti-Horn subjects is a consequence of having fixed all initial priors to favor the frequent object. This means that the model cannot explain initial use of ambiguous *zop* to refer to the infrequent object. This would not be a rational choice for initial pr(infrequent object) $= .25$ neither as τ_c nor as τ_r. However, these subjects' subsequent uses of *zop* tended to be rational provided past evidence, leaving these initial choices unexplained. The remaining error within these first three categories is caused by subjects' occasional experimentation with different signaling patterns. This can be expected in such an experimental setup. Some subjects may take a few trials to explore their options and get accustomed

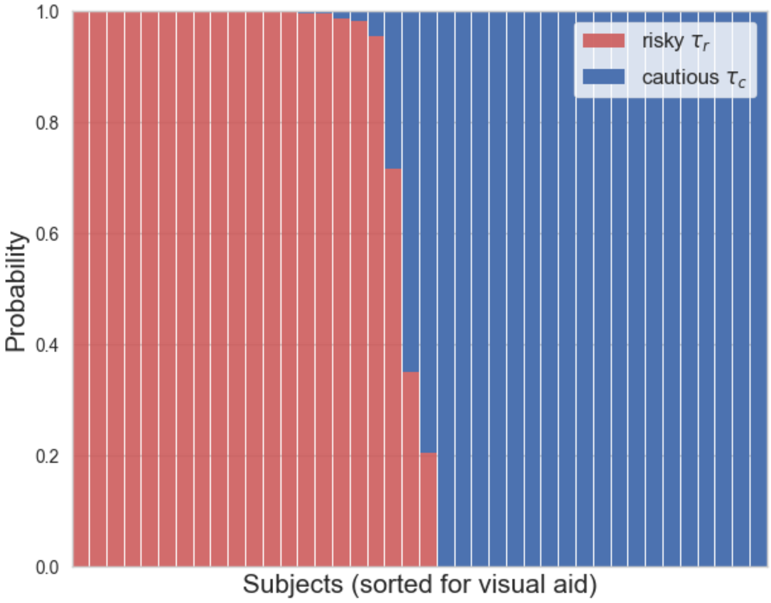


Figure 3.5: Marginal posterior for subjects' types.

to the experiment, as well as to probe their partner's behavior in this unusual task.

As for λ-values, the data is best explained by a tendency toward utility maximization for all but erratic subjects. λ-values are higher for Horn subjects due to the initial prior of $pr(\text{frequent object}) = 0.75$. Such a prior requires high λ for *zop* to be exploited from the onset of the game. That is, before witnessing evidence that this object is actually expected by the receiver.

Figure 3.5 shows the posterior distributions over subjects' types in more detail. There is an almost even split between τ_c and τ_r and uncertainty only about the type of few subjects. This suggests that most subjects' behavior is well explained by either τ_c or τ_r. The few subjects for which some uncertainty remains all belong either to the variational or to the erratic category.

Finally, Figure 3.6 shows draws from the posterior predictive of four different subjects across trials. As discussed below, the bottom-right plot gives an example of a subject which the model cannot explain well. The remaining three plots show cases where the model can explain most of a subject's actual behavior.

3.5.3 Discussion

There are three key things we learn from Kanwal et al.'s (2017) data. First, as illustrated by the left plots in Figure 3.6, different degrees of initial uncertainty about the interlocutor's contextual expectations can explain the behavior of subjects that approximated Horn and unambiguous strategies very well. While it is not possible to ascertain which factors drove subjects to behave as they did,

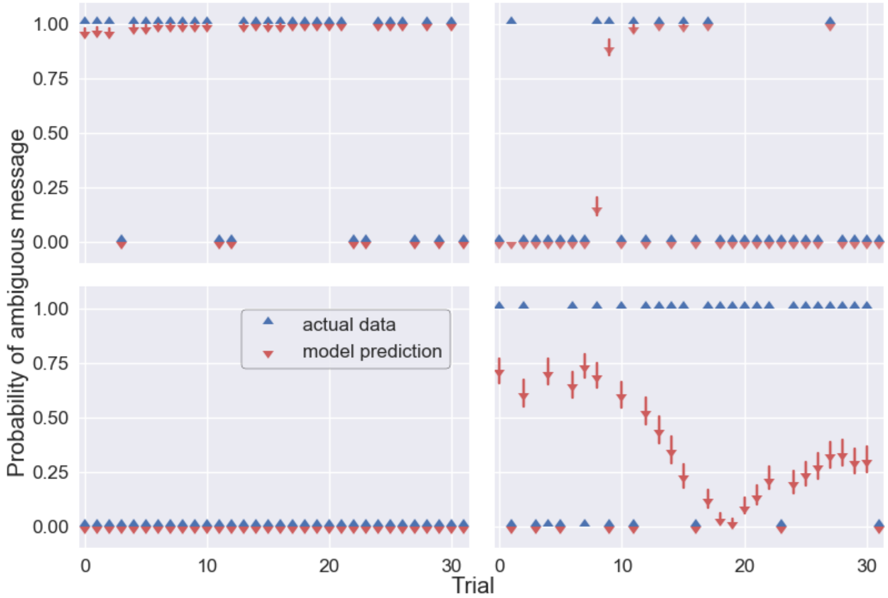


Figure 3.6: Subjects' posterior predictive with 95% confidence intervals, derived from 500 draws per trial. Plots show actual and predicted use of ambiguous *zop* across trials for a Horn subject (top-left), an Anti-Horn subject (top-right), an unambiguous subject (bottom-left), and an erratic subject that experienced 10 communicative failures (bottom-right).

the experiment's exit poll is nevertheless suggestive.[7] In it, many subjects that favored sending long unambiguous labels rationalize their behavior as being motivated by uncertainty, giving answers such as "zop was too easy to be identified as incorrect [I would] rather use the extra time to make sure the guess was correct" or "I wanted to make sure the person got my message. I know it takes time but I prefer to be safe than sorry."

Second, the belief update explains the data well as the game progresses. This is particularly evident in the case of subjects for which the initial prior, which we assumed to favor the frequent object, led to erroneous initial predictions. This is illustrated by the top-right plot in Figure 3.6. The subject's choice not to send the ambiguous message for the frequent object is best explained by assuming her to be of the cautious type τ_c. This, however, predicts the ambiguous message to be avoided across the board. After an initial failure to predict its use for

[7] It is debatable whether subjects' post hoc introspection in this kind of task is informative about their behavior. Additionally, not all subjects provided answers that can easily be interpreted. Some puzzling reasons include "I choose the names depend on the structure of the objects" and "I selected the option which i like the most." Then again, it is to be expected that some subjects will not behave according to theoretical predictions, that some may not be attentive to the task, and that some simply do not care enough about the task to perform it according to the instructions of the experimenters.

the infrequent object, this information feeds into $\mathscr{P}$ and the model catches up with the subject's behavior as she continues using *zop* to signal the infrequent object. This pattern of erroneous initial predictions being improved upon as trials advance holds for all players that approximated the Anti-Horn strategy, as well as for some variational ones.

As noted earlier, it would be possible to improve the fit if we did not fix the initial prior to favor the frequent object. That is, by letting the model estimate an initial prior for each subject. Rather than doing so, the failure on this matter should be seen as indicating need for additional experimentation that controls for subject's expectations in order to understand what factors ultimately lead to the adoption of Anti-Horn-like strategies. Put differently, while the model could explain this data, I do not see a justification within Kanwal et al.'s (2017) experimental setup for the assumption of an initial prior that favors the infrequent object. There may be many reasons why some subjects associated ambiguous *zop* with the infrequent object. However, a better fit would not improve our understanding of this matter. In terms of our overall predictions, this does however suggest that Table 3.4 and Figure 3.5 under-report the amount of risky type τ_r subjects. Currently, Anti-Horn-like subjects are best explained by τ_c even though the posterior predictive suggests this prediction to be off the track. Their use of ambiguous *zop* would instead be best explained by τ_r in combination with an initial prior that favors the infrequent object.

As for variational subjects, their experimentation with multiple strategies is initially unexpected by the model as well. However, once a subject stopped probing her interlocutor's expectations, the model can account for their behavior well.

Some of the subject's comments in the experiment's exit poll suggest reasons for switching strategies. For example, (i) interlocutors' failure to coordinate, "I wanted to set the shorter name for carrots [...] but my partner didn't seem to have understood the game properly"; (ii) starting safe and adapting to the interlocutor's linguistic behavior, "[I w]ait[ed] for the 'partner' to determine what words he/she/it wanted to use [...] and simply [fed] that information back"; or (iii) error "[...] I wasn't thinking well and just used the long names for both, as soon as my partner begin to shorten one I caught on and did the same." The first two reasons fall squarely within the scope of our theoretical analysis. In the first case, learning that the interlocutor is not attentive to the same states as oneself may lead to the adoption of a different strategy. Reversely, positive evidence about an interlocutor's linguistic behavior provides evidence for which strategy best to adopt. The third reason is not accounted for by the model but is also an expected issue that is inherent in empirical data.

Lastly, note that the model fails more the less receptive to communicative success or failure a subject's behavior is. That is, when subjects switched strategies having succeeded in the past with no evidence to support this change. Or, reversely, when they did not switch even in the face of repeated communicative failure. This is illustrated by the bad predictions made for the erratic subject in Figure 3.6 (bottom-right). This subject kept using ambiguous *zop* to signal the frequent object even after her interlocutor had already failed to infer it multiple times, accruing a total of 10 communicative failures. Case in point, after the experiment ended, this subject noted "[...] I always selected zop as the word is small and the time taken [...] would be less." While this is the optimal strategy, the cooperative element of the game seems to have been lost on this subject, who refused to adapt to the partner even in the face of repeated failure. Overall, this issue is only pronounced in a few subjects. The model's failure to predict such behavior is also expected in light of one of our main assumptions being that individuals (at least in tendency) adapt their linguistic behavior to one another. The model fit reflects this by estimating low λ-values for subjects with many communicative failures.

In sum, this analysis further highlights some of the main properties of our model, and some of the aspects in which it differs from previous static treatments such as that of Parikh (2000). Kanwal et al.'s (2017) data is well explained by a tendency toward utility maximization, the possibility of strategies to adaptively change through the course of interactions, and the initialization of $\mathscr{P}$ as effecting either risky or cautious signaling. While, as reported by Kanwal et al. (2017), taken together subjects tend toward an association of ambiguous but preferred forms with frequent meanings, individual-level behavior turned out to foster a multitude of strategies. Where strategies changed the model makes good predictions for subjects that managed to communicate objects successfully. From an experimental perspective, the behavior of Anti-Horn subjects is hard to explain in light of training and testing frequencies. This issue bleeds into the broader question how the prior of individuals and their belief about their interlocutor's prior are formed before interacting (cf. §3.4.2). In the case of Kanwal et al.'s experiment, even if subjects were attuned to frequency differences in training, they had no guarantees that these would transfer to the testing condition, as we assumed, nor that the assigned partner had experienced similar frequencies. This uncertainty may account for some initial variation across subjects. As stressed above, this issue requires further empirical research.

3.6 General Discussion

We proposed a conservative generalization of speaker choice in models of rational language use and combined it with simple adaptive dynamics to generate predictions about ambiguous communication between players lacking a common prior. The model decouples interlocutors' subjective contextual expectations from each other, as well as from the environment itself. This weakens the assumptions of past investigations by neither assuming a common prior (Parikh 2000, Piantadosi et al. 2012b, Santana 2014) nor shared randomness in a language's forms (Juba et al. 2011). Beyond their separation, these components were argued to iteratively feed into each other. A sender's beliefs about her interlocutor play a central role in her linguistic behavior and change according to the receiver's actions. At the same time, interlocutors' communicative intentions and expectations are indirectly shaped by the context and the outcome of interactions. This allows for adaptive behavior that can lead to a plurality of strategies being adopted over time.

In single interactions ambiguity is predicted to be advantageous when (beliefs about) priors are sufficiently aligned relative to the truth-conditions of relevant messages of a language (cf. Parikh 2000, Juba et al. 2011, Piantadosi et al. 2011, Santana 2014). We further showed that these conditions can often be reached when iterated interactions and adaptive mechanisms are considered. Even if players' priors are allowed to initially vary freely. In a nutshell, the more speakers interact, the closer their (beliefs about) contextual expectations grow, and the riskier their communication can be. Crucially, whether (beliefs about) expectations facilitate the safe exploitation of ambiguity is influenced by how informative the context of interaction is. More informative contexts allow interlocutors to reach an implicit agreement on salient meanings faster and more reliably than less informative ones. A byproduct of this interaction is a tendency for the association of preferred forms with frequent meanings (Horn 1984, van Rooy 2004a).

The model also establishes a connection between models of rational language use, usually confined to single interactions, and linguistic alignment. In analogy to experimental findings with human subjects, it predicts increased signal compression as interlocutors interact (Fowler and Housum 1987, Clark and Wilkes-Gibbs 1986, Bard et al. 2000, Motamedi et al. 2016, Kim et al. 2011, Pickering and Ferreira 2008, Brennan and Clark 1996, Hawkins et al. 2017, Kanwal et al. 2017), a strong connection between linguistic adaptation and task success (Fusaroli et al. 2012), and audience and interaction dependent adaptation (Branigan et al. 2010, Garrod and Doherty 1994, Brennan and Clark 1996, Metzing and Brennan 2003). In particular, we showed that the model can explain empirical data in a simple iterated coordination task well. This analysis further underscores that, at least in some contexts, rather than common knowledge of P^* or a common prior, people's behavior in dialog is adaptive and ambiguity exploitation opportunistic. While some subjects tended toward the adoption of the game's optimal strategy, others coordinated on meaning with Pareto-dominated strategies or even unambiguous strategies. This stands in stark contrast to past "one strategy takes it all"-analyses of ambiguity.

The parallels listed above should however not be taken to suggest the model to be a comprehensive model of dialogal adaptation. Our main aim was to add to the general understanding of the conditions under which ambiguity may be justified in cooperative communication, as well as how these conditions can be reached and how they interact. The model is therefore best viewed as an informed but idealized abstraction of communication. It is at this level that it makes predictions about ambiguous communication under the assumption that interlocutors (i) have preferences over messages, (ii) engage in mutual reasoning, (iii) are influenced by information acquired from (iiia) context and (iiib) their interlocutor, and that they (iv) have private contextual expectations. The specifics of these assumptions depend on the situation at hand. For instance, interactions in which linguistic feedback from addressees is limited – such as speeches, lectures or meetings – may require higher degrees of reasoning. Particularly from addressees. On the other extreme, other cases of biological signaling may often involve less rather than more sophistication. In particular, assumptions (ii) and (iiib) may seem contentious when applied to communication of non-human organisms. Along the lines of our and previous accounts, whether ambiguity is explained in functional terms in such cases instead depends on whether priors are generally aligned, dissipating the need for mutual reasoning. An important contributing factor to successful ambiguous communication without conditions (ii) and (iiib) may be that other organisms have been argued to lack or only show very limited degrees of displacement: the ability to communicate about things that are not spatio-temporally present (Hockett 1960). By contrast, in the case of human communication, nothing prevents two zoologists at a baseball court to discuss their work on bats. Taking stock, we proposed a conservative generalization of models of rational language use, embedded it in a dynamic setting in which interlocutors interact multiple times, analyzed its predictions under assumptions that draw from insights of previous research, and explored its explanatory potential using experimental data. Of course, we make no claim to have exhausted the diverse conditions under which biological signaling takes place.

The complementary approach to dialogal adaptation recently proposed by Hawkins et al. (2017) deserves some mention. Rather than starting with fixed semantics, Hawkins et al. analyze adaptation in convention formation. That is, they look at situations where states are yet to be lexically associated with particular forms. The dynamics they propose consequently initialize with interlocutors that are uncertain about the meaning of messages. Technically speaking, interlocutors have uncertainty about their interlocutor's lexicon $L(\cdot, m)$ rather than about her prior (Bergen et al. 2016). Over interactions, evidence for the use of a particular lexicon then leads to the mutual adoption of particular (unambiguous) semantics in a self-reinforcing process initially driven by chance. There are clear parallels and differences between Hawkins et al.'s (2017) and our proposal. In terms of parallels, in both models uncertainty diminishes over interactions and is leveraged to effect agreement on disambiguated language use. As for differences, our prior-driven disambiguation process presupposed fixed semantics. In fact, ambiguous semantics that can be resolved differently across contexts are central to our justification, as well as the starting point of this chapter (see Chapter 5 on the evolution of such semantics). By contrast, lexical uncertainty, as conceived by Hawkins, Bergen, and colleagues, leads to the emergence of unambiguous semantics starting from no preexisting conventions (cf. Skyrms 2010, Spike et al. 2016). I believe that lexical and pragmatic uncertainty are best regarded as dual processes whose explanatory role depends on the degree to which semantics are (believed to be) shared. On the one hand, novel situations may require interlocutors to establish what expressions *mean*. On the other hand, interactions that build on established conventions may instead draw communicative advantage from what expressions can *convey* in a context. I think the rich spectrum of situations where a combination of lexical and pragmatic uncertainty may come into play, as well as a formal and conceptual analysis of their role at the semantics-pragmatics interface offers exciting venues for future research.

One way in which our analysis could be criticized is that players accurately recognize the context they are in and that they approximate subjectively rational behavior (albeit bounded in mutual reasoning depth and allowing for occasional mistakes). These simplifying assumptions do not have a strong bearing on our main argument. A weakening of either is tantamount to the introduction of a higher error rate when using ambiguous signals. It follows that if this rate exceeds the benefit of the use of preferred but ambiguous messages, then unambiguous communication is predicted to be more advantageous and consequently to be adopted. This is well in line with our argument that the benefit of meaning multiplicity is enabled by particular conditions rather than being a property that benefits language users across the board.

This chapter focused on analyzing the conditions under which ambiguous signals can be used without incurring communicative disadvantages in a single context. As noted earlier, one may therefore contend that for any given context an unambiguous language that semantically associates the most frequent state with the most preferred message can be constructed. I agree. Were the world such that language users would always find themselves in exactly the same context there would be little use to associating multiple meanings to a single form. Contextual information would be invariant. Speakers would then do better if they avoided the risk of ambiguous communication altogether and opted for unambiguous expressions instead. It should therefore be stressed that the advantage of expressions that are true of more than one state lies in their ability to fulfill speaker preferences in multiple contexts simultaneously. This is something unambiguous language cannot do. Unambiguous alternatives are nevertheless important; at least for communication that allows for displacement. They come into play either when speakers need to signal a state that is not in line with (beliefs about) contextual expectations, or when these are not sufficiently informative. We will come to further substantiate these claims in Chapter 5.

To summarize, ambiguity endows agents with the ability to adapt their linguistic resources to an environment without incurring too great a risk of misunderstanding. This may involve an adaptation process between interlocutors in a particular situation, but can also draw from general knowledge about commonly experienced domains in single interactions. The more varied the world but more shared the experience, the better ambiguous language users fare. These results add to the growing list of realms in which ambiguity has been argued to be functionally justified, such as non-cooperative communication (Crawford and Sobel 1982), unaligned preferences (De Jaegher and van Rooij 2014), and when a language's form inventory is restricted in size (O'Connor 2015).

3.7 Conclusion

We argued that the risk of ambiguity lies not in the meaning multiplicity of expressions but rather in uncertainty about contextual expectations. In turn, its advantage lies in the reuse of preferred forms, leaving coordination on meaning to be partially resolved by the context of interaction. We have shown under which conditions this justification holds without a common contextual prior and characterized how language users may come to successfully communicate even when these conditions are initially not given, as well as when they fail to materialize. Linguistic alignment was shown to play a pivotal role in this process by having a bearing on coordination and convergence of (beliefs about) expectations over meaning, thereby influencing linguistic choice. In more general terms, we argued that meaning multiplicity is an adaptive tool that enables agents to fit language to their needs, their interlocutors, and the environment, through an exploitation of shared pragmatic principles and (partially) shared contextual information.

Ambiguity is not inevitable. However, when the conditions for its exploitation are given it is likely to emerge through interaction. In functional terms our analysis echoes the sentiment already expressed by Miller (1951:111): ambiguity is not the unruly creature it often is branded to be. Instead, its qualification as disruptive or suboptimal is an artifact of theoretical idealization – a product of expressions' isolated inspection instead of in the naturally richer contexts in which they are produced.

Chapter 4

Co-Evolution of Lexical Meaning and Pragmatic Use

> . . . language is not, as we are led to suppose by the dictionary, the invention of academicians or philologists. Rather, it has been evolved through time, through a long time, by peasants, by fishermen, by hunters, by riders. It did not come from the libraries; it came from the fields, from the sea, from rivers, from night, from the dawn.
>
> Jorge Luis Borges, *This Craft of Verse*

According to standard linguistic theory, the meaning of an utterance is the product of conventional semantic meaning and general pragmatic rules on language use. To investigate how cultural evolution of language plays out under this picture of the semantics-pragmatics interface, this chapter puts forward a game-theoretic model of the competition between types of language users, each endowed with a selection of lexical representations and a particular pragmatic disposition to act on them. The model traces two evolutionary forces and their interaction: (i) functional pressure toward communicative efficiency and (ii) learning biases during the transfer of linguistic knowledge.

We illustrate the model based on a case study on scalar implicatures. In this case study learning biases that favor simple semantic representations are shown to foster the evolution of more sophisticated pragmatic reasoning types and so prevent the lexicalization of scalar implicatures. The picture of the relationship between semantics and pragmatics that this case study suggests is one of co-evolution. The evolution of lexical representations that enable for systematic pragmatic enrichments is dependent on mutual reasoning about rational language use. In the opposite direction, complex semantics that can do away with mutual reasoning do not necessarily foster pragmatic reasoning and may even be encumbered by it.

4.1 Introduction

What is conveyed usually goes beyond what is said. In previous chapters we discussed Grice's (1975) influential characterization of the relation between the literal meaning of expressions and what they may convey in context. Particularly, we saw how the view of pragmatic use and interpretation as a product of mutual reasoning can be captured by models of rational language use (Chapter 2). In Chapter 3 we then analyzed pragmatic inferences that were driven by contextual expectations and beliefs about them. Semantic ambiguity enabled these inferences to be drawn; but their nature is rather ad hoc. Ultimately, they depend on the context and the course of the dialog(s) agents engage in. There are other pragmatic inferences that show striking regularities. For example, the use of ability questions for polite requests ("Could you please ...?"), or certain enrichments of lexical meanings such as that of *and* to convey a temporal succession paraphrasable as *and then.* If we are to understand under which conditions these regular enrichments emerge and stabilize then vertical change needs to be taken into account.

A paradigmatic case of a productive and well studied class of systematic pragmatic enrichments are scalar implicatures (Horn 1984, Hirschberg 1985, Levinson 1983, Geurts 2010; see §2.4). To recapitulate, scalar implicatures refer to inferences where the utterance of a sentence like *I own some of Tom Waits' albums* is taken to convey that the speaker does not own all of them. In the Gricean tradition this inference is viewed as the outcome of the hearer reasoning about the speaker's language use: if the speaker owned all albums, she could have used the word *all* instead of *some*, thereby making a more informative statement. Since she did not, the hearer may conclude that the speaker does not own all of them (under the assumptions that the speaker is cooperative and knowledgeable; that is, willing and able to provide more information if relevant).

Scalar implicatures provide a good testbed to study the evolution of regular pragmatic inferences because they have received much attention, both theoretically (e.g., Sauerland 2004, Chierchia et al. 2012, van Rooij and de Jager 2012) as well as experimentally (e.g., Bott and Noveck 2004, Huang and Snedeker 2009, Grodner et al. 2010, Goodman and Stuhlmüller 2013, Degen and Tanenhaus 2015). While there has been much discussion about many details concerning scalar implicatures, a position endorsed by a clear majority in the literature is that a scalar item like *some* is underspecified to semantically mean *some and maybe all* and that the enrichment to *some but not all* is part of some regular process with roots in pragmatics.

If this majority view is correct, the question arises how such a division of labor between semantics and pragmatics could have evolved, why it would be so pervasive across natural languages, and why it is that some expressions systematically draw from it while others semantically conventionalize.

Models of language evolution abound. There are simulation-based models studying populations of communicating agents (e.g., Hurford 1989, Steels 1995, Lenaerts et al. 2005, Steels and Belpaeme 2005, Baronchelli et al. 2008, Steels 2011, Spike et al. 2016) and there are mathematical models of language evolution, many coming from game theory (e.g., Wärneryd 1993, Blume et al. 1993, Nowak and Krakauer 1999, Huttegger 2007, Skyrms 2010). Much of this work has focused on explaining basic linguistic properties such as compositionality and combinatoriality (e.g., Batali 1998, Nowak and Krakauer 1999, Nowak et al. 2000, Kirby and Hurford 2002, Kirby 2002, Smith et al. 2003, Gong 2007, Kirby et al. 2015, Verhoef et al. 2014, Brochhagen 2015a, Franke 2016), but little attention has been paid to the interaction between conventional meaning and pragmatic use. What is more, many mathematical models explain evolved meaning as a regularity in the overt behavior of agents, abstracting from complex interactions between semantic representations and pragmatic use. In contrast, in this chapter we will look at language users with a richer cognitive make-up. More precisely, we spell out a model of the co-evolution of conventional meaning and pragmatic reasoning. The objects of replication and selection are pairs consisting of a set of lexical meanings and a manner of pragmatic behavior. Put differently, evolutionary forces apply on types of linguistic behavior (Chapter 2), resulting from types' latent semantic and pragmatic properties. This allows us to inspect the influence that evolutionary dynamics have on the joint outcome of particular divisions of labor between semantics and pragmatics.

As in previous chapters, rational language use is modeled using probabilistic models of pragmatic language use (e.g., Frank and Goodman 2012, Franke and Jäger 2016a, Goodman and Frank 2016; see §2.4). Replication and selection are described by the *replicator-mutator dynamic*, a general and established model of evolutionary change in large and homogeneous populations (Hofbauer 1985, Nowak et al. 2000; 2001, Hofbauer and Sigmund 2003, Nowak 2006). This approach allows us to study the interaction between (i) functional pressure toward communicative efficiency and (ii) infidelity in the transmission of linguistic knowledge, caused by factors such as inductive learning biases and sparse learning data. Considering transmission of linguistic knowledge is important because neither semantic meanings nor pragmatic usage patterns are directly observable. Instead, language learners have to infer these unobservables from the observable behavior in which they result. We formalize this process as a form of Bayesian inference. Our approach thereby contains a well-understood model of iterated Bayesian learning (Griffiths and Kalish 2005; 2007), but combines it with functional selection, here formalized as the most versatile dynamic from evolutionary game theory, the replicator dynamic (Taylor and Jonker 1978). Section 4.2 introduces this model.

Section 4.3 discusses the general motivations for the way pressures (i) and (ii) are modeled and combined in light of previous studies. Section 4.4 then applies this model to a case study on scalar implicatures. We discuss a setting in which the majority view of underspecified lexical meanings and systematic pragmatic enrichments emerges if selection and transmission infidelity are combined. In particular, we show that inductive learning biases of Bayesian learners that favor simpler lexical meanings can prevent the lexicalization of scalar inferences and lead to the emergence of Gricean-like pragmatic reasoning types. The results of this case study are critically assessed in the light of the assumptions that feed our model in Section 4.5.

4.2 A Model of Co-Evolving Lexical Representations and Pragmatic Behavior

4.2.1 Communicative success and learnability

The idea that language is an adaptation to serve a communicative function is fundamental to many synchronic and diachronic analyses at least since Zipf's (1949) explanation of word frequency rankings as a result of competing hearer and speaker preferences (e.g., in Martinet 1962, Horn 1984, Jäger and van Rooij 2007, Jäger 2007a, Piantadosi 2014, Kirby et al. 2015). If processes of selection, such as conditional imitation or reinforcement, favor behavior that fosters communicative success, languages are driven toward semantic expressivity (e.g., Nowak and Krakauer 1999, Skyrms 2010; see Chapter 5 for qualification). But pressure toward communicative efficiency is not the only force that shapes language. Learnability is another, as natural languages need to be learnable to survive their faithful transmission across generations. Furthermore, even small learning biases implicit in acquisition can build up and have quite striking effects on an evolving language in a process of iterated learning (Kirby and Hurford 2002, Smith et al. 2003, Kirby et al. 2014).

While natural languages are pressured for both communicative efficiency and learnability, these forces may pull in opposite directions (Christiansen and Chater 2008:§7). Their opposition becomes particularly clear when considering the extreme (Kemp and Regier 2012, Kirby et al. 2015). A language consisting of a single form-meaning association is easy to learn but may fail to convey information agents care about. Conversely, a language that lexicalizes a distinct form for a large number of different meanings may be highly successful in transmitting information faithfully but challenging to acquire.

4.2.2 The replicator-mutator dynamic

An elegant formal approach to capture the interaction between communicative efficiency and learnability is the *replicator-mutator dynamic* (Hofbauer 1985, Nowak et al. 2000; 2001, Hofbauer and Sigmund 2003, Nowak 2006). In its simplest, discrete-time formulation, the RMD defines the frequency x'_i of each type i in an infinite population at the next time step as a function of: (i) the frequency

x_i of each type i before the update step, (ii) the fitness f_i of each type i before the update, and (iii) the probability Q_{ji} that an agent who observes the overt behavior of type j ends up acquiring type i:

$$x'_i = \sum_j Q_{ji} \frac{x_j f_j}{\sum_h x_h f_h} \, . \tag{4.1}$$

The RMD consists of two components: fitness-based selection and transmission perturbations. This becomes most transparent when we consider an equivalent formulation in terms of a step-wise application of the discrete-time replicator dynamic (Taylor and Jonker 1978) on the initial population vector $\vec{x}$ and its subsequent multiplication with a stochastic mutation matrix Q:

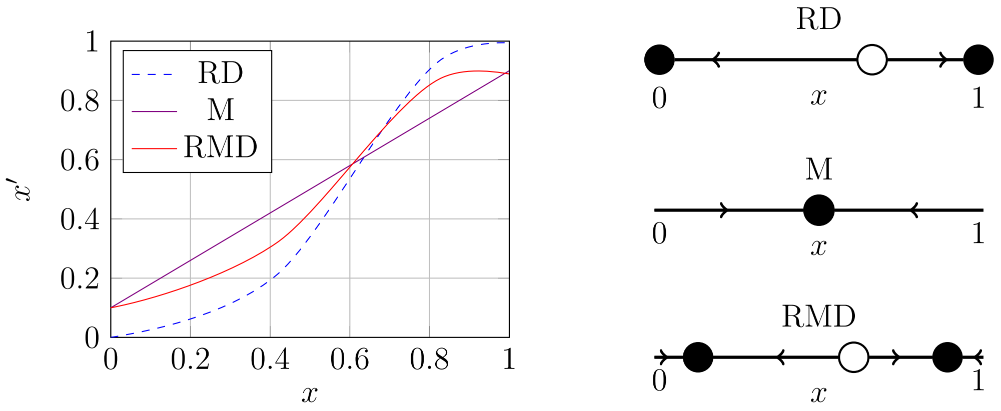
$$x'_i = (\mathrm{M}(\mathrm{RD}(\vec{x})))_i \, , \tag{4.2}$$

where

$$(\mathrm{RD}(\vec{x}))_i = \frac{x_i f_i}{\sum_h x_h f_h} \quad \text{and} \quad (\mathrm{M}(\vec{x}))_i = (\vec{x} \cdot Q)_i = \left(\sum_j x_j Q_{ji} \right)_i \, .$$

The population vector $\vec{x}$ codifies proportions or frequencies of types in a population. This population is infinite; its carrying capacity is held constant, $\sum_i x_i = 1$. This allows us to track change that does not depend on varying population sizes nor their growth rates. Nevertheless, the conceptualizations of replication and mutation with respect to language change suggested below are well compatible with finite populations (see, e.g., Nowak 2006 and Skyrms 2010:§5 for proposals that vary the population, its growth, or the frequency of encounters among types).

If the transmission matrix Q is trivial in the sense that $Q_{ji} = 1$ whenever $j = i$, the dynamic reduces to the replicator dynamic. The replicator dynamic is a model of fitness-based selection in which the relative frequency of type i will increase with a gradient proportional to its average fitness in the population. Selection comes into play whenever two or more types replicate at different rates due to differences in their fitness. As mentioned in Chapter 2, this dynamic is popular and versatile because it can be derived from many abstract processes of biological and cultural transmission and selection (for overview and several derivations see Sandholm 2010). If fitness f_i is the same for all types i, the replicator step is the identity map $(\mathrm{RD}(\vec{x}))_i = x_i$. No difference in fitness translates into a lack of functional pressure. In such cases the dynamic reduces to a process of iteration of the transmission bias encoded in Q: the rate by which one type changes into another after a generational turnover. As detailed below, in this way the process in (4.1), equivalently (4.2), can contain a model of iterated learning (Griffiths and Kalish 2005; 2007). In sum, mutation, encoded in Q and understood as learning in the following, effects selection-independent variation in a population. This variation is maintained or altered through fitness-based selection. The process then repeats.



(a) Update functions: the population state x is mapped onto x' in one update step.

(b) Phase portraits for RD, M and RMD: unstable rest points are hollow, attractors are solid.

Figure 4.1: Example.

Example. Consider a simple and abstract coordination game. Agents are of two types: positive or negative. If agents of different types play with each other, they obtain a payoff of 0. If negative meets negative, each receives a payoff of 1. If positive meets positive, each gets a payoff of 2. We can now inspect how replication, mutation, and their combination affect the composition of populations consisting of agents of these two types.

A population state is completely characterized by the proportion x of negatives. The fitness of negatives in population state x is $f_n(x) = x$. That of positives is $f_p(x) = 2-2x$. The average fitness is $\Phi(x) = xf_n(x)+(1-x)f_p(x) = 3x^2-4x+2$. The replicator dynamic will update x to $\mathrm{RD}(x) = {}^{f_n(x)x}/_{\Phi(x)} = {}^{x^2}/_{\Phi(x)}$. This update function is plotted in Figure 4.1a as the dashed blue line. There are three rest points for which $\mathrm{RD}(x) = x$. These are: $x = 0$, with the population being completely positive; $x = 1$, with the population being completely negative; and $x = {}^2/_3$. The former two points are attractors, meaning that nearby points converge to them. Points near $x = {}^2/_3$ also move toward 0 or 1. This is schematically pictured in the topmost phase portrait in Figure 4.1b. Put differently, there are three cases in which fitness-based selection alone will not change the population's composition. In cases in which the population does not consist of exactly two thirds of negatives it will gravitate toward either extreme rest point.

Adding mutation changes the dynamic and its rest points. For instance, let us assume that $Q_{ji} = .9$ when $j = i$. This is the proportion of types that will retain their type after mutation. Conversely, a proportion of .1 will change their type from positive to negative, and vice-versa. The effects of mutation on its own are described by $\mathrm{M}(x) = .9x + .1(1-x) = .8x + .1$, plotted as the linear violet line in Figure 4.1a. As shown in Figure 4.1b, in this example mutation alone has only one stable rest point. It is located at $x = .5$.

Finally, we can inspect the effects that a combination of replicator and mutator steps have on a population:

$$\text{RMD}(x) = \text{M}(\text{RD}(x)) = \frac{.9x^2 - .2x + .2}{3x^2 - 4x + 2}.$$

This function is plotted in red in Figure 4.1a. The rest points are at $x = .121$, $x = .903$ and $x = .609$. As indicated in Figure 4.1b, the former two are attractors.

4.2.3 Fitness and learnability of lexical meanings and pragmatic strategies

Moving beyond abstract examples, our goal is to apply the RMD to investigate the co-evolution of lexical representations and pragmatic behavior. To do so, we need to fix three things: (i) what the relevant types are, (ii) how fitness derives from communicative success and (iii) how the mutation matrix Q is computed. These issues are addressed, one by one, in the following.

Types: Lexica and pragmatic strategies

Types are what evolution operates on (see §2.2). They define an agent's fitness, usually through a payoff accrued in single interactions with other agents. Often types can be identified as the possible acts in a game; for example, either cooperating or defecting in a prisoner's dilemma. In other cases, they may be thought of as general properties of an agent that influence her fitness, such as being positive or negative in our previous example (whatever that means). For our present purposes, types are identified more concretely by specific assumptions about their cognitive make-up. Since we are interested in the evolutionary competition between different lexical representations and ways of using them in communication, a type is here defined as a pair consisting of a lexicon and a pragmatic strategy of language use. In other words, a type is defined by her reasoning level and the semantic conventions she holds to be true; the sender and receiver behavior resulting from their combination.

As before, a lexicon associates each message with a set of states. A pragmatic behavior specifies a probabilistic sender rule (a probabilistic choice of message for each state) and a probabilistic receiver rule (a probabilistic choice of state for each message) given a lexicon. As discussed in Section 2.4, there are many ways of making these general notions concrete. Here is what we will assume in the remainder of this chapter.

Lexica codify the truth-conditions of expressions. As in Section 2.4, for the case of scalar implicatures we can assume that there are two relevant world states $S = \{s_{\exists\neg\forall}, s_{\forall}\}$ and two relevant messages $M = \{m_{\text{some}}, m_{\text{all}}\}$. For instance, in state $s_{\exists\neg\forall}$ Chris owns some but not all of Tom Waits' albums while in $s_{\forall}$ Chris owns them all. Message m_{some} is short for a sentence like *Chris owns some of Tom Waits' albums* and m_{all} is short for the same sentence with *some* replaced by *all*. Lexica for this case would assign a Boolean truth value to each state-message pair. The following two lexica are minimal examples for the distinction between a lexicalized upper-bound for *some* in L_{bound} and the widely assumed logical semantics with only a lower-bound in L_{lack}.

$$L_{\text{bound}} = \begin{array}{c|cc} & m_{\text{some}} & m_{\text{all}} \\ s_{\exists\neg\forall} & 1 & 0 \\ s_{\forall} & 0 & 1 \end{array} \qquad L_{\text{lack}} = \begin{array}{c|cc} & m_{\text{some}} & m_{\text{all}} \\ s_{\exists\neg\forall} & 1 & 0 \\ s_{\forall} & 1 & 1 \end{array}$$

Pragmatic strategies define dispositions to produce and interpret messages given a lexicon. We distinguish between two kinds of pragmatic strategies. *Literal interlocutors* produce and interpret messages literally, being guided only by their lexica. *Pragmatic interlocutors* instead engage in mutual reasoning to inform their choices. Recall from Section 2.4 that models of rational language use capture different types of linguistic behavior by a reasoning hierarchy. The hierarchy's bottom, level 0, corresponds to literal language use, as in Equations (4.3) and (4.4). Pragmatic language users of level $n+1$ act (approximately) rationally with respect to level-n behavior of their interlocutors, as in Equations (4.5) and (4.6).

$$\rho_0(s \mid m; L) \propto pr(s) L_{[s,m]} \tag{4.3}$$
$$\sigma_0(m \mid s; L) \propto \exp(\lambda\, L_{[s,m]}) \tag{4.4}$$
$$\rho_{n+1}(s|m; L) \propto pr(s)\sigma_n(m|s; L) \tag{4.5}$$
$$\sigma_{n+1}(m|s; L) \propto \exp(\lambda\, \rho_n(s|m; L)) \tag{4.6}$$

According to (4.3), a literal receiver's interpretation of a message depends on her lexicon and her prior over states, $pr \in \Delta(S)$, which is here assumed to be flat (see §2.4 for discussion on the role of priors over states in explaining pragmatic phenomena and of their role for scalar inferences in particular). Literal interpreters thereby choose an arbitrary true interpretation for each message according to their lexicon. Pragmatic receivers, defined in (4.5), instead use Bayes' rule to weigh interpretations based on a conjecture about speaker behavior.[1]

As in previous chapters, sender behavior is regulated by a soft-max parameter $\lambda \geq 0$ (Luce 1959, Sutton and Barto 1998). As λ increases, choices approximate

[1] Note that according to definitions (4.3) and (4.5) hearers do not soft-maximize. This is faithful to the choice we made in Brochhagen et al. (manuscript) and earlier work (Brochhagen et al. 2016). The disadvantage of keeping these definitions unchanged is that only this chapter assumes receiver choice to be belief-oriented (see §2.4 for discussion). This slight deviation is innocuous for this chapter's purposes. The main predictions with or without soft-maximizing receivers do not change; but the particular numeric results do. By keeping the definitions as in previous work, I hope to avoid confusion that might otherwise arise from reporting different numeric predictions across investigations using the same model and case study.

strict maximization of expected utilities. Expected utility of a message m in state s for a level $n+1$ sender is here defined as $\rho_n(s|m;L)$, the probability that the hearer will assign to or choose the correct meaning. For literal senders, utility only tracks truthfulness. Literal senders choose any true message with equal probability but may send false messages as well, with a probability dependent on λ. Differently from the definition of literal sender behavior in (2.14) and (3.2), in this chapter we assume level-0 senders to soft-maximize. The reason is that we want to contrast literal with pragmatic behavior. For this comparison to be feasible, literal behavior needs to be added to the pool of actual strategies that players can adopt, rather than serve solely the purpose of getting pragmatic inference off the ground (see §2.4.2).

The following examples illustrate these behaviors using lexica L_{bound} and L_{lack} from above. A literal interpreter with lexicon L_{bound} assigns $s_{\exists\neg\forall}$ a probability of $\rho_0(s_{\exists\neg\forall} \mid m_{\text{some}}; L_{\text{bound}}) = 1$ after hearing m_{some}, while a literal interpreter with L_{lack} has $\rho_0(s_{\exists\neg\forall} \mid m_{\text{some}}; L_{\text{lack}}) = 0.5$:

$$\rho_0(\cdot \mid \cdot, L_{\text{bound}}) = \begin{array}{ccc} & s_{\exists\neg\forall} & s_{\forall} \\ m_{\text{some}} & 1 & 0 \\ m_{\text{all}} & 0 & 1 \end{array} \qquad \rho_0(\cdot \mid \cdot, L_{\text{lack}}) = \begin{array}{ccc} & s_{\exists\neg\forall} & s_{\forall} \\ m_{\text{some}} & .5 & .5 \\ m_{\text{all}} & 0 & 1 \end{array}$$

By contrast, pragmatic receivers of level 1 have the following interpretative behavior for $\lambda = 1$:

$$\rho_1(\cdot \mid \cdot, L_{\text{bound}}) \approx \begin{array}{ccc} & s_{\exists\neg\forall} & s_{\forall} \\ m_{\text{some}} & .73 & .27 \\ m_{\text{all}} & .27 & .73 \end{array} \qquad \rho_1(\cdot \mid \cdot, L_{\text{lack}}) \approx \begin{array}{ccc} & s_{\exists\neg\forall} & s_{\forall} \\ m_{\text{some}} & .59 & .41 \\ m_{\text{all}} & .35 & .65 \end{array}$$

This is the outcome of reasoning about their level-0 sender counterparts with $\lambda = 1$:

$$\sigma_0(\cdot \mid \cdot, L_{\text{bound}}) \approx \begin{array}{ccc} & m_{\text{some}} & m_{\text{all}} \\ s_{\exists\neg\forall} & .73 & .27 \\ s_{\forall} & .27 & .73 \end{array} \qquad \sigma_0(\cdot \mid \cdot, L_{\text{lack}}) \approx \begin{array}{ccc} & m_{\text{some}} & m_{\text{all}} \\ s_{\exists\neg\forall} & .73 & .27 \\ s_{\forall} & .5 & .5 \end{array}$$

With low λ senders choose true messages with more slack. Reasoning over this behavior therefore also results in a weaker association of messages with only true states in receivers, but also in a slightly stronger association of m_{some} with $s_{\exists\neg\forall}$ over $s_{\forall}$ for L_{lack} users. This is because they reason that $\sigma_0(m_{\text{some}}|s_{\exists\neg\forall}; L_{\text{lack}}) > \sigma_0(m_{\text{some}}|s_{\forall}; L_{\text{lack}})$. For $\lambda = 20$, there will be less slack in literal sender behavior:

$$\sigma_0(\cdot \mid \cdot, L_{\text{bound}}) \approx \begin{array}{ccc} & m_{\text{some}} & m_{\text{all}} \\ s_{\exists\neg\forall} & 1 & 0 \\ s_{\forall} & 0 & 1 \end{array} \qquad \sigma_0(\cdot \mid \cdot, L_{\text{lack}}) \approx \begin{array}{ccc} & m_{\text{some}} & m_{\text{all}} \\ s_{\exists\neg\forall} & 1 & 0 \\ s_{\forall} & .5 & .5 \end{array}$$

And accordingly less slack in level-1 pragmatic interpretation:

$$\rho_1(\cdot \mid \cdot, L_{\text{bound}}) \approx \begin{array}{c} \\ m_{\text{some}} \\ m_{\text{all}} \end{array} \begin{array}{cc} s_{\exists\neg\forall} & s_{\forall} \\ \left[\begin{array}{c} 1 \\ 0 \end{array}\right. & \left.\begin{array}{c} 0 \\ 1 \end{array}\right] \end{array} \qquad \rho_1(\cdot \mid \cdot, L_{\text{lack}}) \approx \begin{array}{c} \\ m_{\text{some}} \\ m_{\text{all}} \end{array} \begin{array}{cc} s_{\exists\neg\forall} & s_{\forall} \\ \left[\begin{array}{c} 0.67 \\ 0 \end{array}\right. & \left.\begin{array}{c} 0.33 \\ 1 \end{array}\right] \end{array}$$

Lastly, turning to types that have no bearing on the choices of receivers of level 1, with $\lambda = 1$ pragmatic senders of level 1 have:

$$\sigma_1(\cdot \mid \cdot, L_{\text{bound}}) \approx \begin{array}{c} \\ s_{\exists\neg\forall} \\ s_{\forall} \end{array} \begin{array}{cc} m_{\text{some}} & m_{\text{all}} \\ \left[\begin{array}{c} .73 \\ .27 \end{array}\right. & \left.\begin{array}{c} .27 \\ .73 \end{array}\right] \end{array} \qquad \sigma_1(\cdot \mid \cdot, L_{\text{lack}}) \approx \begin{array}{c} \\ s_{\exists\neg\forall} \\ s_{\forall} \end{array} \begin{array}{cc} m_{\text{some}} & m_{\text{all}} \\ \left[\begin{array}{c} .62 \\ .38 \end{array}\right. & \left.\begin{array}{c} .38 \\ .62 \end{array}\right] \end{array}$$

For $\lambda = 20$, pragmatic sender behavior of level 1 is instead as follows:

$$\sigma_1(\cdot \mid \cdot, L_{\text{bound}}) \approx \begin{array}{c} \\ s_{\exists\neg\forall} \\ s_{\forall} \end{array} \begin{array}{cc} m_{\text{some}} & m_{\text{all}} \\ \left[\begin{array}{c} 1 \\ 0 \end{array}\right. & \left.\begin{array}{c} 0 \\ 1 \end{array}\right] \end{array} \qquad \sigma_1(\cdot \mid \cdot, L_{\text{lack}}) \approx \begin{array}{c} \\ s_{\exists\neg\forall} \\ s_{\forall} \end{array} \begin{array}{cc} m_{\text{some}} & m_{\text{all}} \\ \left[\begin{array}{c} 1 \\ 0 \end{array}\right. & \left.\begin{array}{c} 0 \\ 1 \end{array}\right] \end{array}$$

There are two particularly important things to note. First, in contrast to their literal counterparts of level 0, pragmatic agents of level 1 using L_{lack} associate m_{some} preferentially with $s_{\exists\neg\forall}$. This association is not perfect, and usually less strong than what agents with a lexicalized upper bound in L_{bound} can achieve – with or without pragmatic reasoning. Higher order reasoning beyond level 1 leads to stronger associations of m_{some} and $s_{\exists\neg\forall}$ also for the receiver. Still, the case study presented in Section 4.4 will consider sender and receiver behavior at levels 0 and 1, as the latter are the simplest pragmatic reasoning types which show a tendency to communicatively attuned pragmatic enrichment. Using only level 1 reasoning and possibly small λ is therefore a conservative choice that works against the fitness-based selection of pragmatic language use for a notion of fitness defined as communicative success, which is introduced next. Second, when it comes to competition between types of use of lexicon L_{bound}, pragmatic reasoning at level 1 is not advantageous. The reason for this is that literal use of L_{bound} already endows agents with a behavioral strategy that associates a single state with a single message (in tendency; depending on λ for senders). For L_{bound}-receivers of level 1, reasoning over stochasticity introduced at σ_0 will generally decrease the association of one state with one message. This decrease is only slight if λ is high, but nevertheless present. That is to say, level-1 reasoning does not necessarily confer a functional advantage. For some types, such as users of L_{bound}, literal signaling is preferable.

Fitness and fitness-based selection based on communicative success

Under the replicator dynamic the proportion of type i in a population will increase or decrease as a function of its relative fitness f_i. In the context of language evolution, fitness is usually associated with the ability to successfully communicate with other language users from the same population (e.g., Nowak and Krakauer 1999, Nowak et al. 2000; 2002). Under a biological interpretation the assumption is that organisms have a higher chance of survival and reproduction if they are able to share and receive useful information via communication with peers. Under a cultural interpretation the picture is that agents themselves strive toward communicative success and therefore occasionally adapt or revise their behavior to achieve higher communicative success (see Benz et al. 2006b:§3.3 and Chapter 2 for discussion).

The replicator equation gives us the means to make the ensuing dynamic precise, without necessarily committing to a biological or cultural interpretation. As above, the proportion of types in a given population is codified in a vector $\vec{x}$, where x_i is the proportion of type i. The fitness of type i is its average expected utility (EU), given the frequencies of types in the current population:

$$f_i = \sum_j x_j \mathrm{EU}(\tau_i, \tau_j) \,. \tag{4.7}$$

In cheap talk signaling games, the expected utility $\mathrm{EU}(\tau_i, \tau_j)$ for type i when communicating with type j is the average communicative success of i when talking or listening to j. If agents are speakers half of the time this yields definition (2.8), repeated below as (4.8).

$$\mathrm{EU}(\tau_i, \tau_j) = {}^1\!/\!_2\, \mathrm{EU}_\sigma(\tau_i, \tau_j) + {}^1\!/\!_2\, \mathrm{EU}_\rho(\tau_i, \tau_j) \,, \tag{4.8}$$

where $\mathrm{EU}_\sigma(\tau_i, \tau_j)$ and $\mathrm{EU}_\rho(\tau_i, \tau_j)$ are the expected utilities for i as a speaker and as a hearer when communicating with j, defined as follows, where n_i and n_j are type i's and type j's pragmatic reasoning types and L_i and L_j are their lexica:

$$\mathrm{EU}_\sigma(\tau_i, \tau_j) = \sum_s P^*(s) \sum_m \sigma_{n_i}(m \mid s; L_i) \sum_{s'} \rho_{n_j}(s' \mid m; L_j) \delta(s, s') \,, \tag{4.9}$$

$$\mathrm{EU}_\rho(\tau_i, \tau_j) = \mathrm{EU}_\sigma(\tau_j, \tau_i) \,. \tag{4.10}$$

As usual, we assume that agents are cooperative, with $\delta(s, s') = 1$ iff $s = s'$ and 0 otherwise (see §2 as well as definitions (2.4) and (2.5), adapted to our present purposes as (4.9) and (4.10), respectively).

Learnability

Languages are shaped not only by functionalist forces toward greater communicative success. While such forces are certainly important for explanations of linguistic change, it is equally important to consider whether other forces might play an explanatory role as well. Another important factor in cultural language evolution is the fidelity with which linguistic knowledge is transmitted. Among others, linguistic production can be prone to errors, states or messages may be perceived incorrectly, and multiple languages may be compatible with the data learners are exposed to. These sources of uncertainty introduce variation in the transmission of linguistic knowledge from one generation to the next. In particular, agents' inductive biases that apply on the iterated transmission process can influence language evolution substantially.

In biological evolution, where types are expressed genetically, transmission infidelity comes into the picture through infrequent and mostly random mutation and genetic drift (Kimura 1983). However, an agent's lexicon and pragmatic reasoning behavior is likely not inherited genetically. They need to be learned from observation. Concretely, when agents attempt to acquire the linguistic behavior of type j, they observe the overt linguistic behavior of type j and need to infer the covert type that most likely produced the observed behavior. As in biology, this transmission process is not perfectly accurate.

Iterated learning is a process in which languages are learned repeatedly from the observation of linguistic behavior of agents who have themselves acquired their behavior from observation and inference. In the simplest case there is a single teacher and a single learner in each generation (e.g., Kirby 2001, Brighton 2002). After sufficient training the learner becomes a teacher and produces behavior that serves as input for a new learner. Figure 4.2a sketches out this idea.

Due to the pressure toward greater learnability it exerts, iterated learning alone generally leads to simpler and more regular languages (see Kirby et al. 2014 and Tamariz and Kirby 2016 for recent surveys). Upon reflection, this arguably intuitive result may give the reader some pause: if iterated learning (at least in tendency) leads to languages with certain properties, for instance, simplicity and regularity, then there must, at some stage in the process, exist factors that would favor languages with these properties over others. Such a factor may be implicit in the way in which the acquisition process is modeled and what feeds it – the kind and quantity of input given to learners and the way in which types are recovered from such input – or it may be an explicit inductive learning bias that skews the learning process toward particular types. The former kind of factor is uncovered by understanding the consequences of the modeling choices taken (e.g., Griffiths and Kalish 2007). One way to make the latter kind of factor explicit and transparent is by modeling agents as Bayesian learners. In this way, inductive learning biases can be encoded in a learning prior over types. Following Griffiths and Kalish (2007) we accordingly model steps in iterated language acquisition as processes of Bayesian inference in which learners combine the likelihood of a type producing the witnessed learning input with prior inductive biases to infer the type that generated this data.

In a Bayesian setting, inductive biases can be codified in a prior over types,

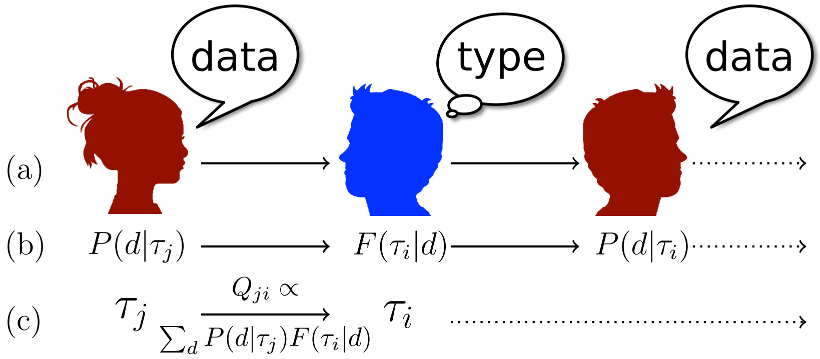


Figure 4.2: Acquisition of type i from a teacher of type j (cf. Figure 1 in Griffiths and Kalish 2007). (a) Naïve learners (left-facing) infer types from overt data, state-message pairs of language use in context, produced by proficient language users (right-facing). The process then repeats: The former learner, now a proficient language user, goes on to produce data for new learners. (b) Dependencies in the process. The probability that type j produces datum d, $P(d|\tau_j)$, is given by her linguistic behavior. The probability of the learner inferring type i from this datum, $F(\tau_i|d)$, combines inductive biases with the likelihood of type i producing the witnessed data. (c) Reduction to transition probability Q_{ji}: the probability that type i is adopted when learning from type j.

$P \in \Delta(T)$, which reflects the amount of data a learner requires to faithfully acquire the type of the teacher (cf. Griffiths and Kalish 2007:450). Put differently, types favored by the prior are easier to learn because they require less data than less favored types to be inferred, even if they are both compatible with the learning data at hand.

The extent of the prior's influence has been shown to heavily depend on the learning strategy assumed to underlie the inference process. On the one hand, early simulation results suggested that weak inductive biases could be magnified by exposing learners to only small data samples (e.g., in Brighton 2002). On the other hand, Griffiths & Kalish's (2005, 2007) mathematical characterization showed that iterated learning alone converges to the prior in the limit. That is, they showed that the resulting distribution over types corresponds to the learners' prior distribution and is not influenced by the amount of learning input given to them. This difference in predictions can be traced back to differences in the selection of hypotheses from the posterior. Griffith & Kalish's convergence to the prior holds for learners that sample from the posterior. That is, for those where learning directly reflects the posterior over types. More deterministic strategies such as the adoption of the type with the highest posterior probability, so-called *maximum a posterior estimation* (MAP), increase the influence of both the prior and the data (Griffiths and Kalish 2007, Kirby et al. 2007). In the following, we use a parameter $\gamma \geq 1$ to modulate between posterior sampling and the

MAP strategy. When $\gamma = 1$ learners sample from the posterior. The learners' propensity to maximize the posterior grows as γ increases.

Let D be the set of possible data that learners may be exposed to. This set D contains all sequences of state-message pairs of length k; for example, $\langle\langle s_1, m_1\rangle, \dots, \langle s_k, m_k\rangle\rangle$. As k increases, learners have more data to base their inference on and so tend to recover the true types that generated a given sequence with higher probability. The mutation matrix Q of the replicator-mutator dynamic in (4.1) can then be defined as follows: Q_{ji} is the probability that a learner acquires type i when learning from an agent of type j. The learner observes length-k sequences d of state-message pairs, but the probability $P(d \mid \tau_j)$ with which sequence $d = \langle\langle s_1, m_1\rangle, \dots, \langle s_k, m_k\rangle\rangle$ is observed depends on type j's linguistic behavior:

$$P(d = \langle\langle s_1, m_1\rangle, \dots, \langle s_k, m_k\rangle\rangle \mid \tau_j) = \prod_{i=1}^{k} \sigma_{n_j}(m_i \mid s_i; L_j)\,, \tag{4.11}$$

where, as before, n_j is j's pragmatic reasoning type and L_j is j's lexicon. For a given observation d, the probability of acquiring type i is $F(\tau_i \mid d)$, so that:

$$Q_{ji} \propto \sum_{d \in D} P(d \mid \tau_j) F(\tau_i \mid d)\,. \tag{4.12}$$

Q is a stochastic $n \times n$ matrix, with n equal to the amount of types. This means that all of Q's cells are numbers in the interval $[0, 1]$ and that each row sums to one. The acquisition probability $F(\tau_i \mid d)$ given datum d is obtained by probability matching, $\gamma = 1$, or a tendency toward choosing the most likely type, $\gamma > 1$, from the posterior distribution $P(\cdot \mid d)$ over types given the data, which is calculated by Bayes' rule:

$$F(\tau_i \mid d) \propto P(\tau_i \mid d)^{\gamma} \text{ and} \tag{4.13}$$

$$P(\tau_i \mid d) \propto P(\tau_i) P(d \mid \tau_i)\,. \tag{4.14}$$

Figures 4.2b and 4.2c summarize this learning process.

4.2.4 Model summary

Communicative success and learnability are central to the cultural evolution of language. These components can be modeled, respectively, as replication based on a measure of fitness in terms of communicative efficiency relative to the population at a given time and iterated Bayesian learning. Their interaction is described by the discrete time replicator-mutator dynamic in (4.1), repeated here:

$$x'_i = \sum_j Q_{ji} \frac{x_j f_j}{\sum_h x_h f_h}\,. \tag{4.15}$$

This equation defines the frequency x'_i of type i at the next time step, based on its frequency x_i before the step, its fitness f_i, and the probability that a learner infers i when observing the behavior of a type-j agent. Fitness-based selection is here thought of not as biological (fitness as expected relative number of offspring) but cultural (fitness as likelihood of being imitated or repeated) evolution, since the types that the dynamic operates on are pairs consisting of a lexicon and a pragmatic use pattern. A type's communicative success depends on how well it communicates within a population while its learnability depends on the fidelity by which it is inferred by new generations of learners. The learners' task is consequently to perform a joint inference over kinds of linguistic behavior and lexical meaning.

The model has three parameters: λ regulates the degree to which senders choose messages that appear optimal from the point of view of the agent's own utility measure (which may be unrelated to the expected utility when communicating with a given population; agents do not change their behavior relative to whom they interact with); k is length of observations for each learner; γ regulates where the learners' inference behavior lies on a spectrum from probability matching to acquisition of the most likely teacher type.

4.3 Functional Pressure: Utility vs. Expressivity

This is not the first model that combines functional pressure with transmission fidelity. Technically closer to our proposal, although applied to problems far remote from the evolution of unobservable interactions at the semantics-pragmatics interface, previous game-theoretic models by Nowak and colleagues have used the RMD in a similar fashion (e.g., Nowak et al. 2001; 2002). These analyses mainly focused on the kind of transmission fidelity necessary for a linguistic type to be adopted by a majority of the population. That is, they addressed the question what value Q_{ii} needs to have for populations to converge on type i; rather than on the question how transmission fidelity and transmission transitions are to be modeled in a principled manner based on types' linguistic behavior. This is where, differently from the work by Nowak and colleagues, (iterated) Bayesian learning comes into play in our application of the RMD.

In the iterated learning tradition much attention has been devoted to the effects that bottleneck sizes and learning biases have on learning alone. Models in this tradition usually lack a communicative element. Language is produced and acquired, but not used to fulfill a communicative task. Functional pressure on successful information transfer consequently plays no role. An exception to this trend is suggested in Kirby et al. 2015. In what follows we discuss aspects of their proposal and contrast it with our own to further motivate our model.

Kirby et al. (2015) propose a model that combines a pressure for learnability with one for *expressivity*. At first sight, it may seem that expressivity and (expected) utility stand for similar concepts. In their own words, "pressure for expressivity arises from language use in communication" (p. 88); it refers to a pressure to be "communicatively functional" (p. 89); and, in fact, it can be "equated [...] with communication" (p. 98). This is also suggested by their laboratory experiment, where this pressure is introduced by having subjects play a signaling game for multiple rounds (Kirby et al. 2015:§3). Their motivation for the introduction of a communicative task draws from previous laboratory experiments using iterated learning. These experiments show that pressure for learnability alone can lead to the emergence of functionally defective languages (e.g., Kirby et al. 2008, Silvey et al. 2014; see Fay and Ellison 2013 for a review of laboratory results). This makes intuitive sense. The simpler form-meaning mappings are, the easier they are to faithfully transmit. Learning a language that maps all meanings to a single form is easier than learning one in which each meaning is associated with a single idiosyncratic form. However, the latter language would generally be better suited to communicate successfully than the former. Pressures that apply on actual communication may be necessary to counteract the drive toward simplicity that learning exerts.

Superficially, it may seem like Kirby et al. have the same pressure in mind as we. However, when inspecting the model's details, expressivity turns out capture a substantially different idea than replication based on communicative fitness. Note first that Kirby et al.'s (2015) model is a pure iterated learning model. In our notation, this means that pressure for expressivity needs to apply within the transmission matrix Q. More precisely, Kirby et al. propose it to affect the production probabilities of teachers that use ambiguous messages. The pressure is regulated by an expressivity parameter. The higher this parameter's value, the more likely a teacher is to produce a false message if messages true of the state she is in are ambiguous. In other words, ambiguous lexica are penalized by making their users more prone to production mistakes. This makes these lexica more difficult to learn because an increased error rate makes it harder for learners to infer the true type that generated this data. While this ambiguity penalty could, in principle, also have consequences for communication with other agents, Kirby et al.'s (2015) model does not have a communicative element involving receivers.

Let us briefly contrast this pressure for expressivity, in the form of inflated error rates when using an ambiguous lexicon, and that for higher utility instantiated by fitness-based replication on a more general level.

Expressivity is absolute and utility relative. Expressivity, as construed above, concerns the ability of speakers of a lexicon to associate each meaning with a single form in production. The degree to which a type is expressive is only determined by the type itself. More precisely, its expressivity is only determined by the type's sender behavior. By contrast, utility concerns the ability to transfer particular information *to someone* (in a particular fashion if talk is not cheap). This is a relative notion that depends not only on the speaker but also on the hearer. Consider, for instance, the two Nash equilibria of the cheap talk 2-states/messages signaling game in Chapter 2. In one equilibrium, m_1 and m_2 are used to signal, respectively, s_1 and s_2. In the other equilibrium, m_1 is associated with s_2 and m_2 with s_1. Senders in either equilibrium are equally expressive. However, a sender following the first equilibrium will always fail to communicate her state to a receiver that followed the second equilibrium, and vice-versa. Similarly, in Chapter 3 we saw that, under certain conditions, ambiguous signaling can be as functionally efficient as unambiguous counterparts, if not more. However, any such advantage is relative to the interlocutor.

Communicative task. Utility makes precise how well a communicative task is fulfilled relative to a population of interlocutors and their communicative preferences. Communicative outcomes are jointly determined by interlocutors that happen to be in this population. Expressivity, in virtue of being absolute, is blind to whether information flows; how well it flows; or to whether some pieces of information are more valuable than others. To give a pointed example for the latter case: if s_1 stands for a state in which I am hungry and s_2 for one in which I am on fire, then I probably care more about signaling s_2 than s_1. If one lexicon allows me to unequivocally convey only s_1 and another only s_2, then I better pick the latter. However, both are equally expressive. Even if we ignore the possibility that there might be different preferences over outcomes, production alone arguably does not adequately capture the task for which language is acquired. Namely, to communicate with other agents in the population.

From the above, we can conclude that expressivity, or a pure learning setting for that matter, does not adequately capture "a pressure to be communicatively functional" (Kirby et al. 2015:89) because it is blind to the task of communicating information to other agents.

Beyond the difference between utility and expressivity there are other smaller differences between our proposal and that of Kirby et al. (2015) in how learning is modeled. However, these differences concern design choices that are well compatible. They do not reflect fundamental differences between the two models.[2] Having clarified the role of functional pressure, we now turn to an application of our model to the case of scalar implicatures.

[2] For instance, in Kirby et al. 2015 naïve learners do not infer a single type, as we have it, but a distribution over types (see also Burkett and Griffiths 2010). When producing data, proeficient language users accordingly first need to sample from their posterior over types and, based on the type that was sampled, produce linguistic evidence for the next learner.

4.4 Case Study: Scalar Implicatures

The model in Section 4.2 formalizes the evolutionary competition between different sets of lexicalizations and ways of using them. This section looks at a case study on scalar implicatures. It engages in a formal thought experiment to address the question: if a population of language users could freely combine different lexica with different pragmatic strategies, what are conditions under which the majority view of scalar implicatures could have evolved?

Recall that the majority view is that scalar implicatures are non-lexicalized pragmatic enrichments. Scalar implicature triggers like *some*, *warm* or *may* are semantically weak expressions for which logically stronger expressions are salient, e.g., *all*, *hot* or *must*. For instance, *some* is entailed by *all*. If the sentence *Chris owns all of Tom Waits' albums* is true, then *Chris owns some of Tom Waits' albums* is also true. However, while weaker expressions such as *some* are truth-conditionally compatible with stronger alternatives such as *all*, this is not what their use is normally taken to convey. Instead, the use of a less informative expression when a more informative one could have been used can license a defeasible inference that stronger alternatives do not hold (cf. Horn 1972, Gazdar 1979). In this way, *Chris owns some of Tom Waits' albums* is strengthened to convey that she owns *some but not all* albums. According to the majority view, this is a pragmatic inference, not part of the conventional meaning.

In the following we consider a specific application of the model from Section 4.2 which allows us to address the question if or when scalar inferences might (not) lexicalize. We consider what is perhaps one of the simplest non-trivial setups that speak to this matter and reflect on its limitations in Section 4.5. The setup is introduced in Section 4.4.1. Section 4.4.2 describes simulations and their results.

4.4.1 Setup

To fill the model from Section 4.2 with life, we need to specify the sets of states, messages, and lexica we consider. Additionally, we want to explore the effects of a learning bias in favor of simple lexical representations. One way of motivating and formalizing such a bias is introduced thereafter.

States, messages, lexical representations, and lexica

Consider a state space with three states $S = \{s_\emptyset, s_{\exists\neg\forall}, s_\forall\}$ and think of it as a partition of possible worlds into cells where none, some or all of the As are Bs, for some arbitrary fixed predicates A and B. Eight lexical representations can be distinguished based on their truth or falsity in three world states, six of which are not contradictory or tautological (see Table 4.2 below).

A lexicon L is a mapping $M \to R$ from messages to representations. With three messages there are $6^3 = 216$ possible lexica. Some assign the same representations to more than one message and others lexicalize the same representations but associate them with different messages. Out of these possible lexica, three kinds are of particular relevance. First, lexica that assign the same lexical representations to more than one message meaning. Such lexica generally lack in communicative efficiency but may be favored by particular learning biases nonetheless (see below). Second, lexica that conventionalize upper-bounds to realize a one-to-one mapping of messages to states. Finally, lexica that do not lexicalize an upper bound but allow it to be conveyed pragmatically due to the presence of a stronger lexical item. There are six lexica of the second kind and six of the third. The following three lexica exemplify each kind:

$$\begin{array}{cc} & \underline{L_{\text{all}}} \\ & \begin{array}{ccc} m_{\text{none}} & m_{\text{some}} & m_{\text{all}} \end{array} \\ \begin{array}{c} s_{\emptyset} \\ s_{\exists\neg\forall} \\ s_{\forall} \end{array} & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \end{array} \quad \begin{array}{c} \underline{L_{\text{bound}}} \\ \begin{array}{ccc} m_{\text{none}} & m_{\text{some}} & m_{\text{all}} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \quad \begin{array}{c} \underline{L_{\text{lack}}} \\ \begin{array}{ccc} m_{\text{none}} & m_{\text{some}} & m_{\text{all}} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \end{array}$$

Lexicon L_{all} is clearly bad for communication: all message and interpretation choices will be equally likely for all reasoning levels; no information about the observed world state will be conveyed by its users. In contrast, users of L_{bound} can communicate world states perfectly, no matter whether they are literal or pragmatic users. Users of L_{lack} can also communicate information about the actual world state but need pragmatic language use to approximate a one-to-one mapping between message use and states (see Section 4.2.3).

Recall that types are a combination of a lexicon and a manner of language use. We analyze the model's predictions in populations of types with one of the two behaviors introduced earlier: literal or pragmatic. The former correspond to level-0 reasoners and the latter to ones of level 1. Accordingly, we consider a total of 432 types. Six are variants of pragmatic language users with L_{lack}-like lexica. We refer to these as *target types* because they represent lexica and language use that conform to the majority view of scalar implicatures. Twelve types are either literal or pragmatic types with lexica of the L_{bound} kind. We refer to these as *competitor types*, because they are expected to be the target types' main contenders in evolutionary competition. Finally, note that while different types may lexicalize the same representations, they may nevertheless map different states to different overt messages. More informally, they speak different languages that lexicalize the same concepts. Consequently, more often than not different lexica of the same kind fail to understand each other (see Section 4.3).

$R \to_2 R \wedge R$	$R \to_2 \neg R$	
$R \to_1 X \subseteq X$	$R \to_1 X \neq \emptyset$	$R \to_1 X = \emptyset$
$X \to_1 \{A, B\}$	$X \to_1 X \cap X$	$X \to_1 X \cup X$

Table 4.1: Toy grammar in a set-theoretic LOT with weighted rules.

intuitive name	$s_{\emptyset}$	$s_{\exists\neg\forall}$	$s_{\forall}$	least complex formula	complexity
"all"	0	0	1	$A \subseteq B$	3
"some but not all"	0	1	0	$A \cap B \neq \emptyset \wedge A \neq \emptyset$	8
"some"	0	1	1	$A \cap B \neq \emptyset$	4
"none"	1	0	0	$A \cap B = \emptyset$	4
"none or all"	1	0	1	$\neg(A \cap B \neq \emptyset \wedge A \neq \emptyset)$	10
"not all"	1	1	0	$\neg(A \subseteq B)$	5

Table 4.2: Available lexical representations and their minimal derivation cost.

An inductive learning bias for semantic simplicity

There is a growing effort to develop empirically testable representational languages that allow for the measure of semantic complexity. For instance, so-called *languages of thought* (LOTs) have been put to test in various rational probabilistic models that show encouraging results (see, e.g., Katz et al. 2008, Piantadosi et al. under review; 2012, and Piantadosi and Jacobs 2016 for recent discussion). At its core, a LOT defines a set of operations and composition rules from which lexical representations can be derived. As a first approximation and for the sake of concreteness, we follow this approach to motivate and formalize a preference of learners for simpler semantic representations (Feldman 2000, Chater and Vitányi 2003, Piantadosi et al. 2012a, Kirby et al. 2015, Piantadosi et al. under review). In a weighed generative LOT a representation's complexity is a function of its derivation cost.

Our toy grammar of lexical representations is given in Table 4.1. This grammar uses basic set-theoretic operations to form expressions which can be evaluated as true or false in states $s_{\emptyset}$, $s_{\exists\neg\forall}$, and $s_{\forall}$ from above. Applications of generative rules have a cost attached to them. Here we simply assume that the formation of Boolean combinations of representations incurs 2 cost units, while all other rule applications incur only 1 cost unit. Table 4.2 lists all six lexical representations relevant here, their truth conditions, and the simplest formula that expresses this representation in the grammar from Table 4.1.

A complexity measure for lexical representations from Table 4.2 is used to define a learning bias that favors simpler representations over more complex ones.

The prior probability of a type is just the prior probability of its lexicon. The prior of a lexicon is a function of the complexity of the lexical representations in its image set. Lexica with simpler representations accordingly have a higher prior. One simple way of defining such priors over lexica (and thereby types) is:

$$P(\tau_i) \propto \prod_{r \in Img(L_i)} P(r), \tag{4.16}$$

$$P(r) \propto \max_{r'} Compl(r') - Compl(r) + 1\,, \tag{4.17}$$

where L_i is type i's lexicon and $Compl(r)$ is the complexity of the minimal derivation cost of representation r according to the LOT-grammar (see Table 4.2). Applied to our space of lexica, this construal assigns the highest probability to a lexicon of type L_{all}, which only uses the simplest lexical representation "all" for all messages. Lexica of type L_{lack} are less likely, but more likely than L_{bound} (see Figure 4.4 below).

There are many ways to define priors over lexica (see, e.g., Goodman et al. 2008, Piantadosi et al. 2012a, Kirby et al. 2015) but the key assumption here, common to all of them, is that simple representational expressions should be favored over more complex ones. We should stress that these details – from the generative grammar to its complexity measure – are to be regarded as one convenient operationalization of one general approach to explicating learning biases; this is not a commitment that this general approach is necessarily superior or that, within it, this particular instrumentalization is the single most plausible.

4.4.2 Simulation results

Recall that there are three parameters: soft-max parameter λ affects how strongly speakers favor messages that appear best from their subjective point of view; the bottleneck size k influences how faithfully learners can identify their teacher type; γ defines the learners' disposition toward choosing the most likely teacher type from the posterior distribution. We expect that competitor types (types with lexica of the kind L_{bound}) have a fitness advantage over target types (pragmatic agents with lexica of the kind L_{lack}), especially for very low levels of λ. Selection based on communicative success alone may therefore not lead to prevalence of target types in the population. On the other hand, lexica of type L_{lack} are simpler than those of type L_{bound} by the postulated measure in (4.16). This may make them more likely to be adopted by learners, especially when k is low so that different teacher types are relatively indistinguishable based on their behavior, and when γ is high. Still, types that use lexica of the kind L_{all} are in turn even more likely a priori than those that use lexica of the kind L_{lack}. Simulation results will shed light on the question whether target types can emerge, and for which parameter constellations.

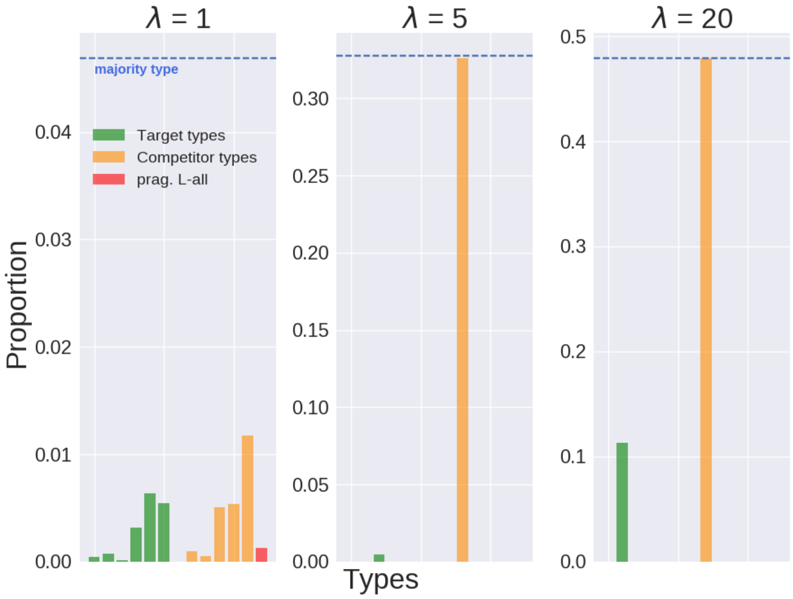


Figure 4.3: Proportion of target types, literal competitor types, pragmatic L_{all}, and the population's majority type, in representative populations after 50 generations under only pressure for communicative success.

To better understand the joint workings of pressure toward communicative efficiency and pressure toward learnability, we look at the behavior of the replicator and mutator step first in isolation, and then in combination. All simulation runs are initialized with an arbitrary distribution over types, constituting a population's first generation. All reported results are the outcome of 50 update steps. These outcomes correspond to developmental plateaus in which change is, if not absent, then at least very slow. In other words, even if the resulting states do not correspond to an eventual attracting state, they characterize almost stationary states in which the system remains for a very long time.

As specified in Section 4.2.3, the mutation matrix Q can be obtained by considering all possible state-message sequences of length k. Given that this is intractable for large k, the sets of data which learners are exposed to are approximated by sampling 250 k-length sequences from each type's production probabilities.

Replication only: selection based on communicative success

Selection based on communicative success is sensitive to λ since this parameter influences signaling behavior. This is showcased in Figure 4.3, which shows the proportion of target types, literal competitor types and pragmatic L_{all} in three representative populations after 50 replicator steps. The plot also indicates the proportion of the *majority type*: the type with the highest proportion in the final population. With low λ many types have very similar behavior, so that evolutionary selection lacks grip and becomes very slow. The result is a very long transition with near stagnancy in a rather homogeneous population with many types. Conversely, higher λ promotes less stochastic linguistic behavior, widening the gap in communicative success between types and promoting more homogeneous populations. As suggested by Figure 4.3, the majority in most populations is not one of the six pragmatic L_{lack}-style types. That is, a pressure only for communicative success does not lead to a prevalence of target types under any λ-value. For instance, with $\lambda = 20$ 1000 independent populations only had 11 cases in which the target type was the majority type, with a mean proportion of .003 across populations. By contrast, in 913 cases the majority type had L_{bound} with close to an even share between literal (454) and pragmatic types (459). Overall, competitor types made up a mean proportion of about .48 taken together.

Iterated learning only: transmission fidelity

The effect of iterated learning without pressure for communicative success using either posterior sampling ($\gamma = 1$) or a stronger tendency toward posterior maximization ($\gamma = 15$) is shown in Figure 4.4 together with the prior over types. The prior shows that while users of L_{lack} are not the most favored by the inductive bias (compared, e.g., to L_{all}) they are nevertheless more advantaged than others, such as L_{bound}. This is due to the relatively simple semantics they conventionalize (see §4.4.1). Crucially, L_{lack} enables its users to convey each state with a single message when combined with pragmatic reasoning and sufficiently high λ. This makes it less likely to be confused with other types if the learning data is not too sparse ($k \geq 5$). Put differently, learners have a higher propensity to infer pragmatic L_{lack} when the teacher's type produces very similar data, such as when using L_{bound}. Moreover, L_{lack} is less likely to be confused with types with different observable behavior because its pragmatic use approximates a one-to-one form-meaning mapping. As a consequence, a stronger propensity to maximize the posterior increases the proportions of targets in the population.

However, in contrast to a pressure only for communicative success with high λ (see Figure 4.3), learnability alone does not succeed in selecting for a single prevalent type. All six target types tend to coexist at roughly equal proportion. Each is passed on to the next generation with the same faithfulness and, differently from a pressure for communicative success, they do not stand in competition with each other (see §4.3). In 1000 independent populations with $\lambda = 20$ and $\gamma = 15$ all majority types were target types, with each reaching approximately the same proportion of users in the population.

As with pressure only for communicative success, low values of λ make the differences in observable behavior across types less pronounced and therefore reflect the learners' inductive bias more faithfully. This favors functionally deficient

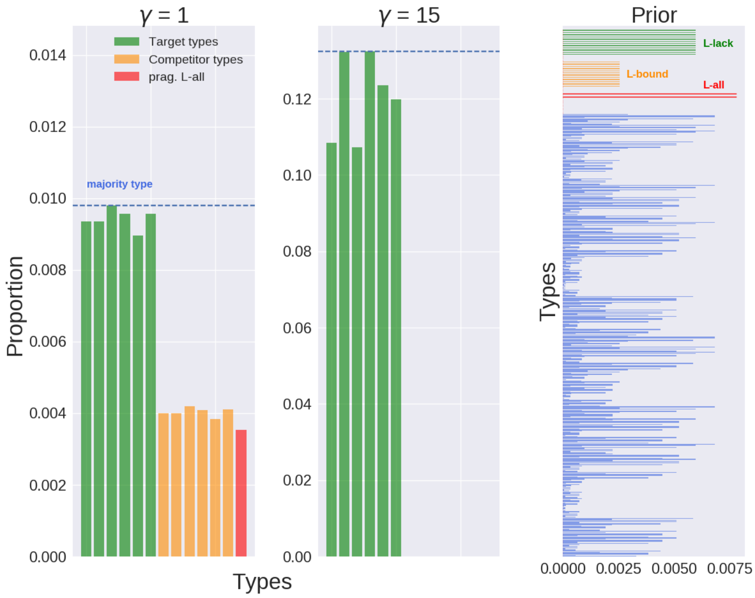


Figure 4.4: Proportion of target types, literal competitor types, pragmatic L_{all}, and the population's majority type in representative populations after 50 generations under only pressure for learnability ($\lambda = 20, k = 5$). The learning prior is shown in the right-most plot with top-most groupings corresponding to types, literal and pragmatic, with lexica of kinds L_{lack}, L_{bound}, and L_{all}.

but a priori preferred types such as those that use L_{all}. As discussed in Section 4.3, pressure for learnability alone can consequently lead to a spread of communicatively suboptimal types that are easier to learn. In the extreme, when $\gamma = 1$ and $\lambda = 1$ all of 1000 independent populations had users of L_{all} as majority types. For higher λ, it becomes clear that a high prior (e.g., L_{all}) is not the only thing that counts for learnability. As soon as there is information for learners to discern whether one type is more likely to have generated the data (depending on λ and k), it becomes paramount for types to produce data that makes them easily identifiable if they are to be transmitted faithfully. This is to say that the learning prior is important but far from being the only element that regulates how iterated learning plays out.

Combining pressures functional pressure and learnability

On its own, a pressure for communicative success or learnability is not sufficient to have a single target type dominate the population. When pressured for communicative success, the communicative advantage of L_{bound} users leads to their prevalence. When pressured for learnability, pragmatic L_{lack} is promoted over functionally similar but semantically more complex alternatives such as L_{bound}. Notwithstanding, learnability alone does not foment the propagation of a single

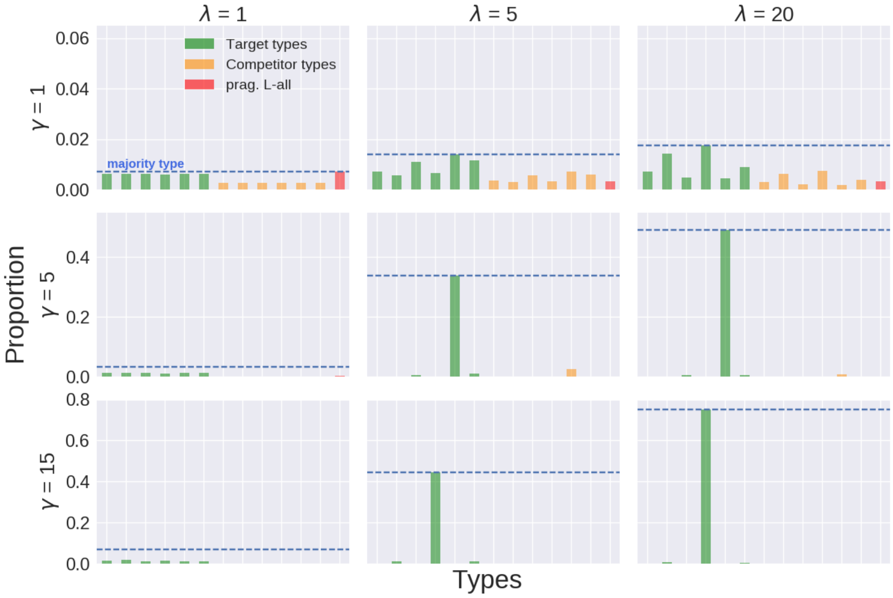


Figure 4.5: Proportion of target types, literal competitor types, pragmatic L_{all}, and the population's majority type in representative populations after 50 generations under both pressures ($k = 5$).

target type across the population.

Figure 4.5 illustrates the combined effects of both pressures for a sample of λ- and γ-values. These results show that an inductive learning bias for simpler semantics in tandem with functional pressure can lead to the selection of a single target type, and so to a division of labor between semantics and pragmatics. The proportion of a single majority target type increases with λ and γ. Pressure for communicative success magnifies the effects of iterated learning and dampens the proliferation of multiple types of a kind that are equal in expressivity *and* learnability (cf. §4.3). A pressure toward learnability favors the transmission of simpler semantics and thereby indirectly promotes pragmatic language use.

As before, low λ and γ lead to the prevalence of communicatively suboptimal types that are a priori favored, such as L_{all}. An increase in λ leads to the selection of target types but does not lead to monomorphic populations if learners sample from the posterior (see the uppermost row in Figure 4.5). Finally, a combination of high λ and γ leads to increasing proportions of a single majority target type. This joint influence is summarized in Figure 4.6, which shows the mean difference between the highest target type and the highest proportion of a type of a different kind in 1000 independent populations across λ and γ values.[3] Higher values of

[3] As before, we cannot just average across target type proportions. This would not be informative about whether just one or multiple target types are present in the population. As illustrated by 4.5 this is particularly important for low λ and γ because multiple types tend to coexist.

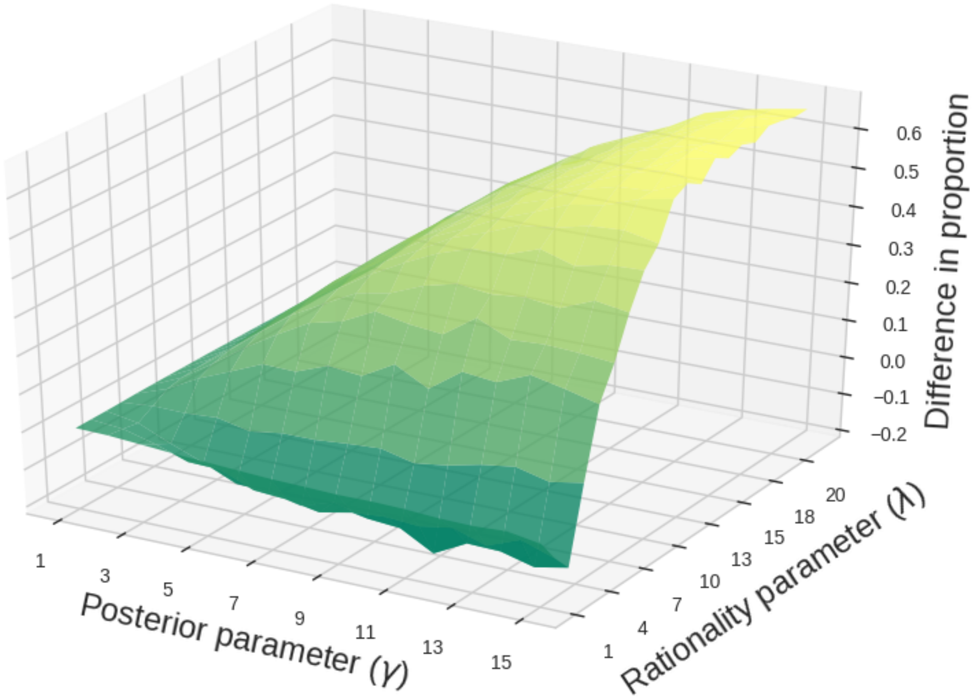


Figure 4.6: Mean difference between proportion of highest target type and highest other type in 1000 independent populations after 50 generations under both pressures across λ- and γ-values ($k = 5$).

λ and γ increase the prevalence of a single target type, whereas lower values lead to less pronounced differences, with a valley resulting from low λ and high γ (cf. Figure 4.5 with $\lambda = 1$ and $\gamma = 15$).

Effects of manipulating the sequence length k have not been addressed so far, but are rather predictable: small values lead to more heterogeneous populations that reflect the learner's prior more faithfully. This is due to the fact that the likelihood that a small sequence was produced by any type is relatively uniform (modulo prior). By contrast, larger values increasingly allow learners to differentiate types with different signaling behaviors.

To recapitulate, other than the involvement of pressure on both communicative success and learnability, the resulting proportion of pragmatic L_{lack} users primarily hinges on three factors. First, the degree, captured by λ, to which agents try to maximize communicative success from their subjective point of view. Second, the inductive bias, which leads learners to prefer simpler over more complex semantic representations in acquisition. Lastly, the learning behavior, captured by parameter γ, where approximating a MAP strategy magnifies the effects of the learning bias in tandem with replication.

More broadly, target types, which represent the majority view of scalar implicatures, can come to dominate the population if three assumptions are met: (i) language is pressured toward both communicative success and learnability; (ii) pragmatic language use is an option; (iii) learners prefer simpler over more complex lexical representations and exhibit a tendency toward the acquisition of the type that best explains the learning data.

4.5 General Discussion

The approach introduced here combines game-theoretic models of functional pressure toward successful communication (e.g., Nowak and Krakauer 1999), effects of transmission perturbations on (iterated) language learning (e.g., Griffiths and Kalish 2007), probabilistic speaker and listener types of varied degrees of pragmatic sophistication (e.g., Frank and Goodman 2012, Franke and Jäger 2014) as well as reasoning about unobservable lexical representations (e.g., Bergen et al. 2012; 2016). This allows for a conceptual investigation of the co-evolution of conventional meaning and pragmatic language use. Main contributions of the model are (i) its modular separation of functional pressure and learnability on evolutionary trajectories, (ii) the characterization of language learning as a joint inference over pragmatic behavior and lexical meaning, and (iii) the possibility to trace the co-evolution of conventional semantics and pragmatic use.

With respect to (i), in Section 4.3 we discussed a comparable model proposed by Kirby et al. (2015) and highlighted the difference between a type's expressivity and functional pressure. As showcased in Section 4.4, the latter pressure can indirectly select for expressive types, i.e., those that can convey states unequivocally. By contrast, Kirby et al.'s model only considers the bearing that expressivity has on the production of learnable data. We see three main reasons for considering utility rather than just expressivity; and, relatedly, to consider the effects that communication *and* learning have on an evolving linguistic system, rather than only learning. First, learning alone can promote populations with non-negligible proportions of functionally defective types. This is true both of simulations, e.g. our L_{all}-types in Figure 4.4, as well as of laboratory experiments with human subjects (see §4.3 and references within). Second and more importantly, types may be equally expressive but their performance as means of information transfer crucially depend not only on themselves but on the population they find themselves in. That is, we contend that adopting an expressive type that generates learnable data does not in itself capture a type's arguably central communicative function of transferring information to peers. Taking communication into consideration allows the model to be responsive to the task for which language is learned. Lastly, chains of iterated learning alone do not put types in direct competition. Accordingly, learning can lead to polymorphous populations in which multiple variants of a type coexist (compare the competition of target types in Figure 4.5 and their lack of competition in Figure 4.4). This latter point of course depends on the particulars of a model's learning component. Here, we explored the effects that simple transmission chains have, but it is possible to add further complexity to this process. For example, by letting learners learn from multiple teachers (Burkett and Griffiths 2010, Brochhagen et al. 2016).

The main result of our case study is that types that correspond to the majority view of scalar implicatures – that scalar readings are non-lexicalized pragmatic enrichments – can come to dominate a population. This can happen under the assumption that simpler semantic representations are more likely to be learned (Chater and Vitányi 2003). Pragmatic language use can be recruited indirectly by such preference for simpler lexical representations. Under this view, semantics and pragmatics play a synergic role: pragmatic use allows maintenance of simpler representations; pressure toward representational simplicity indirectly promotes pragmatic over literal language use. As a consequence, iterated transmission and use of language lead to a regularization that may explain the lack of lexicalization of systematic pragmatic enrichments.

While the results of this case study are interesting, they also raise a number of critical issues. First of all, while many favorable parameter settings exist which lead to a prevalence of target types, other types are usually represented in non-negligible proportions. This may just be a technical quirk of the mutator step. But there is a related issue of empirical importance. Several experimental studies on scalar implicatures suggest that participants can be classified as either semantic or pragmatic users of, in particular, *some* (e.g., Bott and Noveck 2004, Nieuwland et al. 2010, Degen and Tanenhaus 2015). The former consistently accept *some* where *all* would be true as well, the latter do not. Interestingly, in our simulations when a target type is the majority type, an inflated proportion of the population uses compatible lexica with a lexicalized upper bound. Particularly in those parameter settings where the prior influences the outcome less. In other words, we do find a tendency toward a similar co-existence of semantic and pragmatic types. Whether this analogy has any further explanatory value is an interesting path for future exploration.

Another important issue that is not addressed in the model are potential costs associated with pragmatic reasoning. Here, we simply assumed that literal and pragmatic reasoning strategies exist from the start and are equally costly to apply. In contrast, empirical results suggest that the computation of a scalar implicature may involve additional cognitive effort (e.g., Breheny et al. 2006, De Neys and Schaeken 2007, Huang and Snedeker 2009, Tomlinson Jr. et al. 2013; but see also Foppolo and Marelli 2017 for evidence to the contrary). Extensions of the model presented here to include processing costs for pragmatic language use would be straightforward but interesting future work. It seems plausible that effects of reasoning cost may trade off with the frequency with which a given scalar expression is used (see Chapter 5 on the effects that frequency can have). It may be that frequently drawn scalar implicatures lexicalize to avoid cost, whereas infrequent ones are derived on-line to avoid more complex lexical representations during acquisition. Such a prediction would lend itself to empirical testing in line with recent interest in differences between various scalar implicature triggers (van Tiel et al. 2014).

Our case study could be criticized as follows: all it shows is that scalar implicatures do not lexicalize because upper bounds are dispreferred lexical representations; the result is just due to learning. This criticism would be too superficial and highly unjust. Dispreferred lexical representations can thrive under evolutionary selection. Lexicalized upper-bounds can dominate a population because they may boost communicative success. But they do not have to. Moreover, even without selective pressure for communicative success, it is not necessarily the case that the types that are most likely a priori will dominate. The dynamics of iterated learning are not that trivial. Iterated learning does not necessarily promote the a priori more likely type, but tends to promote a type i based on a gradient of how many other types might likely mutate into i, so to speak. Taken together, without an explicit model of the interaction between pressure for communicative success and learnability, it is far from trivial to judge whether, or when, preferred or dispreferred representations evolve (see §2.3 where we argued for dynamic over static analyses this reason, among others, and Chapter 6 for general discussion on this matter). This is why a major contribution of this chapter is the arrangement of many different ingredients into a joined model of the co-evolution of lexical meaning and pragmatic use.

What is more, it is not that we just assumed a prior disadvantage of lexicalized upper bounds. We tried to motivate and formalize a general assumption about lexical representations' complexity with a concrete, albeit provisional proposal. The specification of a learning bias in terms of a "grammar of representations" can and should be seen critically, however. Much depends on the primitives of such a grammar. For instance, the lexical representation "none or all" is the most complex in Table 4.2. But consider adding a new primitive relation between sets $A \smile B$ which is true if and only if $\neg(A \cap B \neq \emptyset \wedge A \neq \emptyset)$. The lexical representation "none or all" would then be one of the simplest. Clearly, further research, empirical and conceptual, into the role of representational complexity, processing costs and learning biases is needed. The model here makes a clear and important contribution nonetheless: it demonstrates how simplicity of representations can interact with use and evolutionary selection and shows that for simple representations to emerge it may require pragmatic strategies to compensate their potential functional deficiencies. Hence a model of co-evolving semantics and pragmatics is needed. Future work should also include the possibility that representational simplicity may itself be a notion that is subject to evolutionary pressure (cf. Thompson et al. 2016), as well for the evolution of elements that define the agents' cognitive make-up: λ and γ.

Finally, this case study should not be interpreted as a proposal for a definite explanation of how scalar implicatures evolved. Other factors should be considered eventually even if they will lead to much more complex modeling. One such factor is the observation that non-lexicalized upper bounds allow a broader range of applicability. For example, when the speaker is not certain as to whether *all* is true. This may suggest an alternative and purely functionalist argument for why upper-bounded meanings do not conventionalize: should contextual cues provide enough information to the hearer to identify whether a bound is intended to be conveyed pragmatically, then this is preferred over expressing it overtly through longer expressions. For example, by saying *some but not all* explicitly. Importantly, although morphosyntactic disambiguation may be dispreferred due to its relative length and complexity (Piantadosi et al. 2012b), it allows speakers to enforce an upper-bound and override contextual cues that might otherwise mislead the hearer. In a nutshell, this explanation posits that scalar implicatures fail to lexicalize because, all else being equal, speakers prefer to communicate as economically as possible and pragmatic reasoning enables them to do so (cf. Chapter 3). What this alternative argument does not explain is why functional pressure does not lead to the emergence of different, equally costly lexical items to express different knowledge states of the speaker (Horn 1984:252-267, Horn 1972, Traugott 2004, van der Auwera 2010). For instance, to the emergence of two expressions for each weak scalar expression; one with and one lacking an upper-bound. That is, it does not explain why English and other languages do not have a monomorphemic dual expression for, e.g., *some* that lexicalizes an upper-bound. If this hypothetical expression existed, it could be deployed to signal that the speaker knows that *some but not all* holds, and unbounded *some* could exclusively signal epistemic uncertainty. Looking at pressure from learnability might come in again.

Beyond scalar implicatures, the model can generate predictions about likely lexicalization trajectories of pragmatic inferences, or a lack thereof. In this realm an interesting issue is whether proposed principles, such as the semantic conventionalization of once highly context-dependent inferences if they become regular enough (Levinson 2000, Traugott 2004), can be given a formal rationale and inform postulated directionalities of change. The present chapter made a first start and gave a framework for exploring these issues systematically.

4.6 Conclusion

The cultural evolution of meaning is influenced by intertwined pressures. We set out to investigate this process by putting forward a model that combines a pressure toward successful information transfer with perturbations that may arise in the transmission of linguistic knowledge in acquisition. Its objects of selection and replication are pairs of lexical meanings and patterns of language use. This allows the model to trace the interaction between conventional meaning and pragmatic use. Additionally, it takes the challenge seriously of neither semantics nor pragmatics being directly observable. Instead, learners need to infer these unobservables from overt data that results from their combination. These components and their mutual interaction were highlighted in a case study on the lack of lexical upper-bounds in weak scalar expressions that showed that, when pressured for learnability and communicative success, the former force drives for simpler semantic representations inasmuch as pragmatics can compensate for functional disadvantages in use. That is, the relative learning advantage of simpler semantics in combination with functional pressure in use may offer an answer to why natural languages fail to lexicalize certain systematic pragmatic inferences. And, more broadly, lead to a division of labor between semantics and pragmatics.

Chapter 5

Evolution of Ambiguity

> Each city receives its form from the desert it opposes.
>
> Italo Calvino, *Invisible Cities*

In Chapter 3 we showed that semantic ambiguity can be functionally advantageous provided that interlocutors' (beliefs about) contextual information agrees, leading to successful disambiguation; or that they interact multiple times so that speakers come to anticipate receivers' interpretative behavior when faced with ambiguous signals. The extent to which this advantage crystallizes was shown to depend on the context(s) in which interaction takes places. More precisely, on the objective distributions over states that govern these contexts. A central assumption that this analysis built on was that conventionalized form-meaning associations enabled for the exploitation of ambiguity in the first place. That is, we assumed that interlocutors based their signaling behavior on lexica in which preferred messages are semantically associated with two or more states. The present chapter seeks to fill the gap covered by this assumption by elucidating whether and when conventional semantic meaning that enables for functional ambiguity exploitation evolves. We do so by considering not only horizontal but also vertical change, using the model from Chapter 4.

Our results suggest that semantic ambiguity can indeed evolve if there is functional pressure for efficient information transfer and pressure for learnability. However, this only happens if the world is such that communication and learning take place in a mixture of contexts, each governed by a different state distribution. In particular, for ambiguous semantics to survive their faithful transmission across generations, communication needs to take place in informative contexts in which different states are frequent. This is necessary for naïve learners to receive sufficient evidence that a signal is semantically associated with multiple states. Should communication instead take place in a homogeneous world in which only a single state is frequent, or in one in which no state is frequent relative to others, then unambiguous semantics conventionalize.

5.1 The Evolution of Ambiguity: A Puzzle to be Explained?

As mentioned in Chapter 3, much work on the emergence and stability of linguistic conventions has focused on conditions under which *unambiguous* signaling emerges (e.g., Lewis 1969, Steels 1998, Skyrms 2010; see Spike et al. 2016 for a recent review). By contrast, investigations of pragmatic inference in terms of rational language use standardly take as their starting point some kind of semantic underspecification (e.g., Franke and Jäger 2016a, Goodman and Frank 2016); or they consider other factors that introduce uncertainty over meaning. For instance, a noisy channel (Bergen and Goodman 2015). After all, a shared one-to-one form-meaning mapping in an environment that allows for noiseless communication leaves little to no room for pragmatic refinement.

How can these two strains of research be consolidated? One the one hand, language use can come to exploit ambiguity through pragmatic reasoning. On the other hand, work on language emergence tells us that the association of multiple states with a single message is "bad news" (Skyrms 2010:68); at least when communicative success hinges on distinguishing these states, and players have at least as many messages available as there are states.

On a general level, this apparent contrast is easy to dispel. Work on the emergence and transmission of language usually explains evolved meaning as a regularity in the overt behavior of agents, abstracting from complex interactions between semantic conventions and pragmatic use. That is, a distinction between semantics and pragmatics is seldom, if ever, drawn. This means that this line of research should not be viewed as explaining regularities in underlying relationships of form and semantic meaning, but rather as explaining regularities in the overt linguistic behavior of members of a population (Lewis 1969); call them signaling strategies or pragmatic language use. Once an interaction between semantic meaning and factors that influence how it is deployed in context are factored in, the bad news about ambiguity need to be qualified: an outcome is suboptimal only if language use, operating over semantic meaning, gives rise to uncertainty over states. Under this view, the apparent tension between these two strains of research disappears.[1]

In short, ambiguous signaling behavior, but not necessarily ambiguity at the semantic level, is functionally disadvantageous and puzzling from an evolutionary perspective. In Chapter 3 we surveyed many functional advantages ambiguity can confer, such as smaller vocabularies; greater signal compression; reuse of preferred

[1] Of course, if no information beyond conventional semantic meaning is at play then semantics is directly reflected by overt signaling behavior. One may argue this to be true of particular natural language phenomena, or of certain cases of non-human signaling. However, it should be relatively uncontroversial to argue for a distinction between semantics and pragmatics where contextual information or mutual reasoning are involved, as in context-driven disambiguation (Chapter 3) or in certain pragmatic inferences (Chapter 4).

forms that are easier to produce or parse; or coordination on novel meaning, for example, in the form of metaphors. What we then want to understand is under which conditions ambiguous semantic conventions evolve and stabilize provided that actual signaling behavior can sometimes turn ambiguity to its advantage.

In light of the synchronic functional advantages of ambiguity, the main challenge we face concerns vertical change. This challenge can be framed as follows. Linguistic knowledge needs to survive its faithful transmission across generations, being iteratively passed on to naïve learners. These learners need to infer unobservables, such as semantic meaning, from the overt linguistic behavior of their teachers. If patterns of language use are not to be functionally disfavored, they have to exhibit (a tendency toward) unambiguous signaling in a given context. This gives rise to the following tension: it is functionally advantageous to signal unambiguously but this can disadvantage the acquisition of ambiguous semantics since the overt behavior that learners witness may not suggest underlying ambiguity.

This chapter's goal is twofold. First, we want to complement our analysis of signaling with ambiguous messages from Chapter 3 by elucidating under which conditions lexica that allow for functional ambiguity exploitation evolve. Second, we want to explore the predictions of our model from Chapter 4 by applying it on a different question; looking at a different type space, as well as a different inductive bias; and to analyze the influence that communication and learning in different contexts have on language evolution at the semantics-pragmatics interface.

Section 5.2 summarizes the model from Chapter 4 and introduces the setup we focus on. Section 5.3 shows our main results. We discuss them in Section 5.4 and conclude in Section 5.5.

5.2 Model Summary and Setup

As before, we model the interaction between functional pressure and learnability using the replicator-mutator dynamic (Hofbauer 1985, Nowak et al. 2000; 2001, Hofbauer and Sigmund 2003, Nowak 2006). The discrete RMD, defined in (4.1) and repeated below, describes change in an infinite population $\vec{x}$ as a function of (i) the frequency x_i of each type i before the update, (ii) the fitness f_i of each type i, and (iii) the probability Q_{ji} that a learner witnessing overt behavior of type j will end up with type i (see Section 4.2 for details and discussion).

$$x'_i = \sum_j Q_{ji} \frac{x_j f_j}{\sum_h x_h f_h} \, . \tag{5.1}$$

The fitness of type i is defined as its expected utility in the population. Intuitively, fitness indicates how well a type communicates with members of her community. The transmission matrix Q codifies transition probabilities. These give the fidelity with which a type is passed on to the next generation of signalers. Q_{ji} is the probability of type i being acquired when learning from type j.

Learning is here defined as a process of (iterated) Bayesian learning in which a learner infers a type from the observable behavior of her parent/teacher (Griffiths and Kalish 2005; 2007). The value of Q_{ji} depends on two factors. First, it depends on the probability $P(d \mid \tau_j)$ of witnessing datum d when learning from type j. This is the likelihood that a teacher of type j produces particular messages when in particular state. Second, Q_{ji} also depends on the probability $F(\tau_i \mid d)$ that the learner infers witnessed datum d to have been generated by type i: $F(\tau_i \mid d) \propto P(\tau_i \mid d)^\gamma$ where $P(\tau_i \mid d) \propto P(\tau_i)P(d \mid \tau_i)$. The learning prior, $P(\tau_i) \in \Delta(T)$, codifies inductive biases that the learner may bring to the task. The combination of the prior with the likelihood of type i producing d yields the learner's posterior. The posterior, in turn, is regulated by parameter $\gamma \geq 1$ which controls whether learners sample from it, $F(\tau_i \mid d)^1 = P(\tau_i \mid d)$, or whether they instead have a tendency to maximize the posterior. This tendency grows as γ increases.

On the one hand, a fitness differential between types leads to the selection of fitter types. In linguistic terms, this amounts to a pressure for successful and efficient communication. On the other hand, if $Q_{ji} \neq 1$ for $j = i$, then the transmission of linguistic knowledge from one generation to the next is perturbed. This can have striking effects on an evolving linguistic system. In particular, if the faithfulness to which a type is passed on depends on its learnability, as assumed here, then types are also pressured for being inferable from overt and possibly sparse linguistic input.

The fitness of a type depends on the company it keeps and the context(s) of interaction in which communication takes place. A type may be well equipped to communicate with some types but may fail to do so when interacting with others. Moreover, it may be better equipped to communicate some states than others. If the distribution over states that governs a context (dis)favors certain states, then this may also affect a type's fitness. In Chapter 4 we tacitly considered a single and uniform objective distribution over states. In this chapter, we analyze how variation in a distribution over state distributions, i.e., variation in the frequency in which agents find themselves in different contexts, can affect the evolution of ambiguous semantics. We first introduce this idea and its consequences in general terms. The type space we inspect by simulation is introduced afterward.

5.2.1 Contexts and objective state distributions

Our analysis of ambiguity in Chapter 3 showed that the functional (dis)advantage semantic ambiguity confers depends on the context of interaction and the distribution over states P^* that governs it. In the extreme, if there are two states s_1 and s_2 but $P^*(s_1) = 1$ then it is irrelevant whether an ambiguous but preferred message could be used to signal state s_2. Senders would never find themselves in this state. Drawing from Chapter 3, we want to inspect how the contexts in which communication takes place influence the kind of semantic conventions a population adopts.

There are two straightforward ways in which the influence of state frequencies could be inspected in detail. The first is to consider a single distribution P^* that changes over time. The second is to consider a distribution over state distributions, $\mathscr{C} \in \Delta(P^*)$, which regulates the frequency in which agents find themselves in a context governed by a particular P^*. The second alternative is what we assume in the following. It has the advantage of not having to define a rate of contextual change and additionally allows us to easily inspect how the frequency in which communication happens in certain contexts affects the evolution of ambiguity.

In terms of fitness, we simply need to add $\mathscr{C}$ to the computation of expected utility (cf. definitions (4.9) and (4.10) for unique P^*).[2] For discrete $\mathscr{C}$ and S, the expected utility of type i communicating with type j as a speaker, $\mathrm{EU}_\sigma(\tau_i, \tau_j)$, and as a hearer, $\mathrm{EU}_\rho(\tau_i, \tau_j)$, are defined as follows:

$$\begin{aligned}\mathrm{EU}_\sigma(\tau_i, \tau_j) = & \sum_{P^*} \mathscr{C}(P^*) \sum_s P^*(s) \sum_m \sigma_{n_i}(m \mid s; \mathscr{P}; L_i) \\ & \qquad \sum_{s'} \rho_{n_j}(s' \mid m; pr^j; L_j) \; (\delta(s, s') - c_\sigma(m))\,;\end{aligned} \tag{5.2}$$

$$\begin{aligned}\mathrm{EU}_\rho(\tau_i, \tau_j) = & \sum_{P^*} \mathscr{C}(P^*) \sum_s P^*(s) \sum_m \sigma_{n_j}(m \mid s; \mathscr{P}; L_j) \\ & \qquad \sum_{s'} \rho_{n_i}(s' \mid m; pr^i; L_i) \; \delta(s, s'),\end{aligned} \tag{5.3}$$

where, as before, n_i and n_j are type i's and type j's pragmatic reasoning types, L_i and L_j are their lexica, pr^i and pr^j are their subjective priors over states, and $\mathscr{P}$ is the sender's belief about the receiver's prior over states (see below for a review on how signaling behavior is defined). As before, fitness is defined as:

$$f_i = \sum_j x_j \mathrm{EU}(\tau_i, \tau_j)\,, \tag{5.4}$$

where

$$\mathrm{EU}(\tau_i, \tau_j) = {}^1\!/_2\, \mathrm{EU}_\sigma(\tau_i, \tau_j) + {}^1\!/_2\, \mathrm{EU}_\rho(\tau_i, \tau_j)\,. \tag{5.5}$$

In terms of learning, we will assume that learners are aware of the context in which the linguistic input they receive is produced. To this end, we need to

[2] Differently from Chapter 4 but following the motivations for ambiguous signaling given in Chapter 3, messages are assumed to carry some cost for senders in this chapter.

distinguish between a context of interaction and the distribution over states that governs this context. The former is what learners witness, together with the state and message produced in this context. More precisely, where before the learners' input was k-length sequences of $\langle s, m\rangle$-pairs, now learners observe sequences of indexed $\langle s, m\rangle_c$-pairs with $c \in C$ being the context in which m was observed to signal s. In short, language use is situated in context and learners are aware of this.

The true distribution in context c is P^*_c, but P^*_c itself is not accessible to learners. They are only able to distinguish one context from another. The distribution over state distributions $\mathscr{C}$ nevertheless has an impact on learning as it affects the data teachers produce. For k-length datum $d = \langle \langle s_1, m_1\rangle_1, \dots, \langle s_k, m_k, \rangle_k\rangle$ we now have that:

$$P(d \mid \tau_j) = \prod_{i=1}^{k} \mathscr{C}(P^*_i) \;\; P^*_i(s_i) \;\; \sigma_{n_j}(m_i \mid s_i; \mathscr{P}; L_j)\,, \tag{5.6}$$

where, as before, n_j is j's pragmatic reasoning type and L_j is j's lexicon.

To illustrate the effect that the existence of multiple contexts of language use has on learning, consider a situation with two states, s_1 and s_2, three contexts, u, v, and w, and their respective distributions $P^*_u(s_1) = .9$, $P^*_v(s_1) = 1$ and $P^*_w(s_2) = 1$. Let there be only two types, i and j. Both always use message m to signal s_1. However, one uses m' and the other uses m'' to signal s_2, $m' \neq m''$. If $\mathscr{C}(P^*_v) = 1$ then the data they produce will be indistinguishable. Message m is not informative about whether i or j generated the learning input and all that learners witness in this case are sequences of observations of the form $\langle s_1, m\rangle_v$. Less extremely, if $\mathscr{C}(P^*_u) = 1$, then some data sequences may contain messages uttered in s_2. These messages can tease i and j apart. The linguistic input that learners receive will be even more informative if $\mathscr{C}(P^*_w) > .1$.

The same issue arises for types that use an ambiguous message to signal different states in different contexts. If $\mathscr{C}$ is degenerate, their overt linguistic behavior will be indistinguishable from that of a type that uses an unambiguous lexicon. Intuitively, if *bat* is used to refer to baseball bats in a sports context but to refer to animals in a zoo, there will be little evidence for the ambiguity of *bat* if communication happens almost exclusively in one of the two contexts.

In sum, the existence of multiple contexts that differ in state frequency not only affects whether or to which degree players find themselves in a context that may be (dis)favorable to their type, communication-wise. It also affects the data that learners witness. Both of these factors are central to the issue at hand given that (i) a functional advantage for semantic ambiguity depends on the distribution over states and consequently it also depends on the frequency in which agents find themselves in a context (Chapter 3), and that (ii) for semantic ambiguity to be faithfully transmitted, overt language use in context needs to suggest that a message is conventionally associated with multiple states.

5.2.2 Type space

As before, a type is a combination of a lexicon and a disposition to use it to communicate in context. In the following, the latter will correspond to the level-1 behavior of (boundedly) rational language users as they were defined in Chapter 3. We repeat the relevant definitions for player i below.

$$\rho^0(s \mid m; pr^i; L) \propto L_{[s,m]}\, pr^i(s); \tag{5.7}$$

$$\sigma^0(m \mid s; L) \propto L_{[s,m]} - c_\sigma(m); \tag{5.8}$$

$$\rho^1(s \mid m; pr^i; L) \propto \exp(\lambda \frac{\sigma^0(m \mid s; L) pr^i(s)}{\sum_{s'} \sigma^0(m \mid s'; L) pr^i(s')}); \tag{5.9}$$

$$\sigma^1(m \mid s; \mathscr{P}; L) \propto \exp(\lambda((\int \mathscr{P}(\theta)\rho^0(s \mid m; \theta; L)d\theta) - c_\sigma(m))). \tag{5.10}$$

Recall that the level-1 receiver defined in (5.9) results from reasoning about level-0 sender behavior in (5.8). Receiver i's tendency to maximize utility from her own perspective grows as λ increases and pr^i is her subjective prior over states. Such a receiver takes senders to signal following the semantic conventions she holds true, codified in her lexicon L_i, and combines this behavioral expectation with her contextual expectations, codified in $pr^i \in \Delta(S)$.

The level-1 sender in (5.10) is our generalization of rational sender behavior. This sender reasons about literal level-0 receiver behavior, as defined in (5.7). This reasoning process involves the sender's beliefs about her addressee's prior over states $\mathscr{P}$, with priors parametrized in θ. Intuitively, if in the context of interaction a rational sender believes her addressee to expect state s_1 rather than state s_2, then she may attempt to exploit this expectation by sending a preferred ambiguous message in s_1, but not in s_2. As with receiver behavior, λ regulates the strength of the sender's tendency to maximize utility from her subjective perspective.

In Chapter 3 we saw that, over time, simple adaptive dynamics can lead interlocutors' priors over states to converge to the true distribution P^* that governs a context. In the following, we relax the assumption of a non-common prior to allow for a more succinct analysis, abstracting away from proximate causes that lead interlocutors that share a set of semantic conventions to coordinate on ambiguous signals in informative contexts. We assume all types' priors to correspond to P^*.

Following our setup in Chapter 3, sender i's beliefs about her interlocutor's prior, $\mathscr{P}$, are Dirichlet distributed with weights for state s set to $q \times pr^i(s) + 1$. As q increases, so does the sender's belief that the receiver's prior is close to her own. In this setup, this is equivalent to the belief that the receiver's prior is close to true P^*. On the lower end, $q = 0$ corresponds to full uncertainty about the receiver's prior.

For the simulations in Section 5.3 we assume λ and q to be common as well. The reason for these simplifying assumptions is that we want to trace change in types' lexica and the effects the context of interaction has on the evolution of ambiguous semantics, rather than to consider a situation where competition hinges on variation in priors over states, q-values, λ-values, or reasoning levels. Accordingly, and in contrast to Chapter 4, we assume all types to be level-1 reasoners. Type i is therefore fully determined by her lexicon L_i.

Turning to the space of lexica that we consider, recall that in Chapter 3 we had a lexicon that specified the truth-conditions of three messages for two states, with $c_\sigma(m_1) = .4 = c_\sigma(m_2)$ and $c_\sigma(m_3) = .1$ as message cost. Following this setup, we consider a space of lexica that is made up of all possible state-message mappings for this 2-states/3-messages game that lexicalize no contradictory message.

We will think of messages as being of the form *This is an x*, with a different x in each state (see below for details on how this changes how the learners' inductive bias is conceptualized and further motivation). This yields three possible message meanings: either $s_1 \vee s_2$, if the message is true in both states; $s_1 \wedge \neg s_2$, if true in s_1 but false in s_2; or $\neg s_1 \wedge s_2$, if false in s_1 but true in s_2. With three messages there are $3^3 = 27$ lexica, and accordingly 27 types in our type space.

The restriction to non-contradictory messages does not imply that every message is necessarily employed. To see this, consider the following two lexica:

$$L_t = \begin{array}{c} \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{ccc} m_1 & m_2 & m_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{array} \qquad L_{a2} = \begin{array}{c} \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{ccc} m_1 & m_2 & m_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{array}$$

Lexicon L_t is one of our *target* lexica. It associates preferred message m_3 with both s_1 and s_2 but also lexicalizes unambiguous messages to signal these states. Lexicon L_{a2} exemplifies a lexicon with two ambiguous messages: m_3 and m_2. Assume that senders strongly believe their interlocutors to expect s_2. For instance, that it is believed with certainty that $pr(s_2) = .9$. In this case, rational users of both lexica alike would use preferred m_3 when in state s_2, and unequivocal m_1 when in state s_1. Message m_2 is not used in s_2 because it is more costly than m_3. Crucially, the fact that m_2 is ambiguous in L_{a2} but unambiguous in L_t does not need to lead to a difference in their overt signaling behaviors. Whether there is an observable contrast between certain types will depend on the frequency in which they find themselves in contexts that lead them to behave in different ways. That is, it will depend on the distribution over state distributions $\mathscr{C}$.

Our analysis will focus on two kinds of types. The first use lexica with ambiguous m_3 and unambiguous m_1 and m_2, as exemplified by L_t above. We call these target types because their lexica correspond to the ones we analyzed in Chapter 3: they enable for the use of preferred m_3 in both states, but also lexicalize unambiguous alternatives that are employed when uncertain about their interlocutors' contextual expectations. The second kind are unambiguous *competitor types* that do not lexicalize ambiguous messages. Lexicon L_u, below, illustrates such a lexicon. There are two target types and six competitor types in total.

$$L_u = \begin{array}{c c} & \begin{array}{ccc} m_1 & m_2 & m_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \end{array} & \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \end{array} \qquad L_{a3} = \begin{array}{c c} & \begin{array}{ccc} m_1 & m_2 & m_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \end{array} & \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{array}$$

Relative to the entire type space, target types are rather conservative when it comes to how much lexical ambiguity they harbor. They are more ambiguous than competitor types. However, other lexica, such as L_{a2} and L_{a3}, lexicalize more ambiguous messages.

The functional competition between targets and competitors is relatively straightforward. Target types can exhibit more flexible signaling behavior by using preferred m_3 in both states. However, in uninformative contexts they might opt for more costly m_1 and m_2. By contrast, competitor types have a safe unambiguous signaling strategy from the get-to, but are disadvantaged against targets if $\mathscr{C}$ favors a mixture of distributions over states with both frequent s_1 and s_2. As in Chapter 4, target types and other users of ambiguous lexica will tend to exhibit more stochastic behavior than competitors. Particularly if λ is low.

5.2.3 Inductive learning bias

In Chapter 4 learners had to infer the semantic meaning of quantifiers such as *all* or *some*. These meanings were expressed by formulae to explore the effects that an inductive bias that favors simple semantic representations over more complex ones has. In the following we conceptualize the relevant meanings to be inferred from messages as object extensions. The reason for this shift, beyond the fact that lexical ambiguity is a natural and intuitive form of ambiguity, is that it allows us to explore the consequences of a learning constraint often associated with the difficulty of learning ambiguous lexical labels: the mutual exclusivity bias (Markman and Wachtel 1988, Merriman et al. 1989, Clark 2009). In other words, we want to see whether ambiguous semantics of the target type can evolve if they are a priori dispreferred by learners over those that competitor types lexicalize, and do so motivated by a well studied acquisition bias.

Mutual exclusivity refers to a learning constraint that plays an important role in the acquisition of novel linguistic labels. Markman and Wachtel (1988) famously registered it in a series of experiments. For instance, in their first experiment they instructed 3-year-olds to "Show me the l" where l was a novel linguistic label. Children had to pick between two objects: one with a name they already knew and one that they did not know the name of. For example, children had to decide whether l referred to a banana (known name) or a lemon wedge press (unknown name). Overall, Markman and Wachtel found that children show a strong tendency to infer that the novel label applies to the object they do not know the name of. This tendency has been taken to suggest a learning bias for linguistic labels to be mutually exclusive.

Following Markman and Wachtel's study, mutual exclusivity has attracted much attention. To name a couple of details under active investigation, there seem to be age differences in how strong this bias is (e.g., Halberda 2003, Bion et al. 2013); it has been suggested that it extends well into adulthood (e.g., Halberda 2006); and mutual exclusivity seems to be stronger in monolingual learners than in plurilingual ones (e.g., Bialystok et al. 2010). These findings have led some to argue the bias to be shaped by a learner's linguistic experience, being more of a malleable word learning strategy than a fixed preference (Houston-Price et al. 2010).

Merriman et al. (1989) suggest a number of functions for mutual exclusivity. For instance, it may aid learners' word learning process by serving as a heuristic to map labels to objects. On this front its effect is akin to the pragmatic strengthening that results from mutual reasoning about rational language use: if the speaker wanted to refer to the object with the known name she would have used the known label, since she did not, the name must apply to the unlabeled object. Mutual exclusivity may also aid in reorganizing and correcting the semantic conventions a learner entertains. For example, the extension of a known word, say *dog*, may be corrected upon learning a novel label for an object assumed to fall under its extension, say *wolf*. Without such a bias, the hypothesis that *dog* also applies to wolves would remain intact.

The mutual exclusivity bias is evidently not absolute. Children and adults alike do learn and use near-synonyms, such as *leaves* and *foliage*, and words below and above the so-called basic-level, e.g., not only *dog* but also *dalmatian* and *animal*. More generally, they come to master multiple ways to refer to an object, be it *baseball bat*, *baseball club*, *bat*, *club*, or *thing*.

We implement mutual exclusivity as a learning prior that favors lexica that do not map multiple messages to a single object. For $L \in \{0,1\}^{|S|\times|M|}$ and writing $L_{[s*]}$ for the row in L corresponding to s:

$$P(L) \propto \exp(|S| - b \sum_{s \in S} \text{count}(L_{[s*]}); \tag{5.11}$$

$$\text{count}(L_{[s*]}) = \begin{cases} \sum\limits_{m \in M} L_{[s,m]} - 1 & \text{if } \sum\limits_{m \in M} L_{[s,m]} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5.12}$$

Parameter $b \in [0;1]$ regulates the strength of the mutual exclusivity bias. Since types differ from one another only in terms of their lexica, the prior probability of a type is that assigned to its lexicon: $P(\tau_i) = P(L_i)$. If $b = 0$ the prior is flat. For $b > 0$ mutual exclusivity leads to the distinction of four kinds of lexica in our type space. From most to least favored these are: (i) lexica that associate only one state with two messages, such as the competitor lexicon L_u, (ii) those that associate two states with two messages each, such as target L_t, (iii) those that

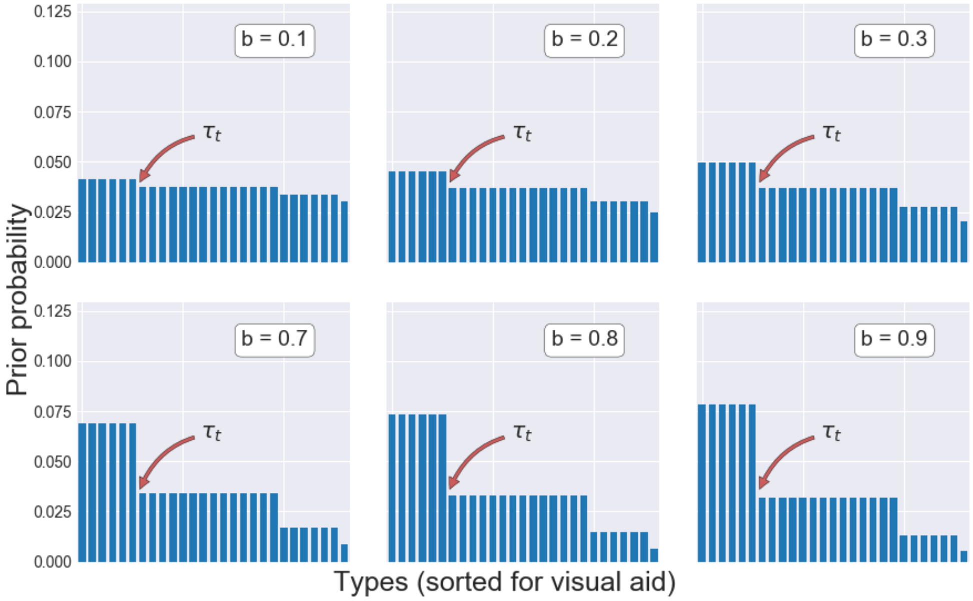


Figure 5.1: Prior over types for different values of parameter b. The arrow shows the group to which target types belong when types are ordered from most to least favored.

associate one state with three messages and one with two, such as L_{a2}, and (iv) fully ambiguous lexica that associate all messages with all states, as L_{a3}.

Figure 5.1 shows the prior for different values of b. As can be read off from these plots, if the prior is not flat then six types fall within the most favored category; 14 fall within the second-most favored category; six into the third-most favored category; and fully ambiguous L_{a3} is alone in the least favored category.

5.2.4 Summary

Our goal is to see whether, and if so under which conditions, ambiguous semantics evolve. Drawing from previous chapters, we focus on the effects that functional pressure and pressure for learnability have, relative to the frequency in which agents find themselves in a context. In particular, we focus on how these factors influence the evolution of ambiguous lexica of the target type, exemplified by L_t above.

The frequency by which players find themselves in context c_i, governed by $P_i^* \in \Delta(S)$, is controlled by $\mathscr{C}$. We expect the main contenders of target types to be users of unambiguous lexica of the L_u-kind. First, because learners prefer the latter lexica a priori in virtue of associating less messages with the same state. Second, if $\mathscr{C}$ favors either only a (close to) degenerate context, a (close to) uniform one, or a mixture between these two, then L_u-style lexica can be functionally advantageous: they do not depend on pragmatics to disambiguate ambiguous signals, nor is there a functional advantage to associating multiple states with a single message if either (i) contexts do not allow for safe ambiguity exploitation, or (ii) they always favor the same state with high probability (Chapter 3).

5.3 Simulation Results

Our setup involves six parameters: q regulates the degree to which senders believe the receiver's prior over states to be close to theirs; λ regulates how strongly senders/receivers favor messages/interpretations that appear best from their subjective point of view; sequence length k influences how much input learners receive and consequently how faithfully they can identify their teacher's type (relative to how closely the teacher's overt behavior resembles that of other types in the population); γ modulates the strength of learners' tendency to maximize the posterior; b controls the strength of the mutual exclusivity bias; and $\mathscr{C}$ is the distribution over state distributions which determines how frequently agents find themselves in a particular context.

As in Chapter 4, we begin by inspecting functional pressure and pressure for learnability in isolation in order to gain a better understanding of their effects on this type space. We focus mainly on the influence of $\mathscr{C}$ over that of other parameters. As detailed below, $\mathscr{C}$ regulates much of the types' competition and transmissibility. Once its influence is factored in, the effect of the remaining parameters are consistent with the trends reported in Chapter 3 and 4.

Each population is randomly initialized. All reported simulations correspond to population states after 500 update steps. These outcomes correspond to developmental plateaus in which change is, if not absent, then at least very slow. As before, computing Q for large k is intractable. We therefore approximate the mutation matrix by sampling 1000 k-length sequences from each type's production probabilities. For expository ease, we consider only three distributions over states. These are $P^*_1(s_1) = .9$, $P^*_2(s_1) = .1$, and $P^*_3(s_1) = .5$. In words, in context c_1 state s_1 is much more frequent than s_2. Context c_2 reversely favors state s_2. Context c_3 is uniform. To keep our notation simple, we write $\mathscr{C}(P^*_i)$ as $\mathscr{C}_i$.

Figure 5.2 contains the space that $\mathscr{C}$ spans. As exemplified by the five circular nodes in this figure, we will explore our model's predictions at the points of the 3-simplex in which $\mathscr{C}_1 + \mathscr{C}_2 + \mathscr{C}_3 = 1$.

Drawing from the preceding discussion, we expect that a distribution over state distributions with either high $\mathscr{C}_1$, or high $\mathscr{C}_2$, or a mixture of only one of the former with $\mathscr{C}_3$ will not lead to a prevalence of target types. In such cases, competitor types using L_u-style lexica will be as – if not more – functionally advantageous while being easier to learn. By contrast, and in accordance to our analysis in Chapter 3, we expect a distribution $\mathscr{C}$ that spreads its probability across contrasting state distributions, P^*_1 and P^*_2, somewhat evenly to be the most conducive for target types, at least when it comes to fitness-relative replication.

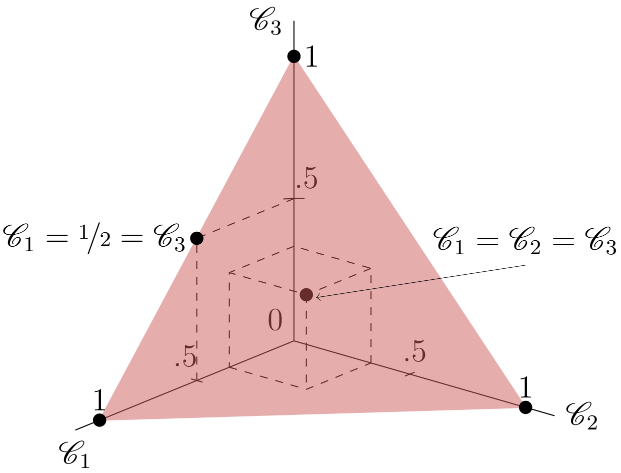

$\mathscr{C}_3$
1
.5
$\mathscr{C}_1 = 1/2 = \mathscr{C}_3$
$\mathscr{C}_1 = \mathscr{C}_2 = \mathscr{C}_3$
0
.5
.5
1
1
$\mathscr{C}_2$
$\mathscr{C}_1$

Figure 5.2: The standard 3-simplex. Edge values range from 0 to 1. $\mathscr{C}$ is degenerate at the three vertices with a node. It is uniform at the simplex' center.

5.3.1 Functional pressure only

Players' signaling behavior depends on q and λ. The value of q, however, only affects senders of types with lexica that allow for the exploitation of ambiguity but also lexicalize unambiguous alternatives. In particular, with sufficiently high λ, low q leads target senders to avoid using preferred m_3 because they do not believe ambiguous signals to be safe. Instead, if fueled by large enough q they will use preferred but ambiguous m_3 in salient states (see Chapter 3). For the remainder, we fix $q = 40$ to the effect that target types have a tendency to send m_3 in s_1 if in c_1; in s_2 if in c_2; and to avoid ambiguity if in uniform c_3. This enables us to investigate under which conditions exploitable ambiguity of the target kind emerges.

The influence of $\mathscr{C}$ and λ on expected utility is straightforward. In a world in which $\mathscr{C}_1$ and $\mathscr{C}_2$ are high but $\mathscr{C}_3$ is low, target types have a functional advantage over other types when communicating with other agents of equal type. This difference in expected utility becomes more pronounced the higher λ is. For instance, for $\lambda = 1$ and $\mathscr{C}_1 = .45 = \mathscr{C}_2$, $\text{EU}(\tau_i, \tau_i)$ for all types ranges from approximately .5 to .55. The two target types and the six competitor types all come close to the latter value. For the same $\mathscr{C}$ but $\lambda = 20$ the expected utility of type i communicating with others of its type ranges from approximately .76 to .92. Under these circumstances, the two target types alone have the highest expected utility when communicating with their own type, trailed by competitor types. If $\mathscr{C}$ favors either only context c_1 or context c_2 then target types do as well as competitors for $\lambda > 10$ but worse for lower values. For example, if $\mathscr{C}_1 = .9$ and $\mathscr{C}_2 = \mathscr{C}_3$. Finally, in a world in which c_3, governed by a uniform distribution over states, is more frequent than either c_1 or c_2, target types lose their functional advantage. Part of the reason for this is that if players find themselves more often in uniform c_3 than in c_1 or c_2 then using m_3 to signal exclusively either one of the states is better than to avoid its use altogether. Moreover, as mentioned above, the use of L_t-style lexica carries a risk of misunderstanding that types with unambiguous m_3 do not suffer from. If $\mathscr{C}$ favors contexts where only s_1 or only s_2 is frequent, then it is more advantageous to have a lexicon that unequivocally associates m_3 with the frequent state.

Taking stock, there are two central things to note in terms of expected utility. First, higher λ leads to a starker contrast between types. This is not only true of $\text{EU}(\tau_i, \tau_i)$ nor particular to this type space, but is a more general consequence of the rationality parameter λ. Low values promote stochastic behavior that blurs differences that some types would exhibit if they had a stronger tendency toward expected utility maximization. Second, as aforementioned, whether target types have a functional advantage depends on $\mathscr{C}$. If $\mathscr{C}_1$ and $\mathscr{C}_2$ are both high, L_t-style lexica are particularly advantageous. Conversely, competitors and other types that use m_3 only in a single state do better than target types if $\mathscr{C}_3$ is higher than at least either $\mathscr{C}_1$ or $\mathscr{C}_2$; or if a single context is much more frequent than others. This makes intuitive sense. The functional advantage that the exploitation of m_3 in both s_1 *and* s_2 can confer does not come to bear its fruits if the world is such that players only communicate either s_1 or s_2, or if the context is uninformative and ambiguity is avoided.

Inspecting only expected utility, and more so only a fragment of it, can be misleading. After all, fitness and replication depend on the population agents find themselves in. Figure 5.3 shows (i) the mean difference between the highest proportion of target types and the highest other type in 1000 independent populations, as well as (ii) the mean difference between the highest proportion of competitor types and the highest other type for $\lambda \in \{1, 5, 20\}$ across values of $\mathscr{C}$ (see Figure 5.2). This figure shows that functional pressure alone promotes target types only in the small region in which both $\mathscr{C}_1$ and $\mathscr{C}_2$ are high, and $\mathscr{C}_3$ is very low; and only if λ is high. The converse is true of competitor types, who only thrive when λ is low and the environment leads to frequent communication in uniform c_3. As for the remaining types, none of them comes close to establishing itself in the population under these parameter constellations.

Overall, these results suggest that, for most values of $\mathscr{C}$, the functional advantage of target ambiguous lexica is not strong enough to promote either variant of this type. The outcomes in which competitor types come to dominate should also be seen critically. They result from leveraging the erratic signaling behavior effected in other types by low λ.

5.3.2 Learnability only

As may be intuited from Figure 5.1, which shows the learners' prior, if $0 < b < 1$ then its particular value only has a slight impact on differences in the learnability

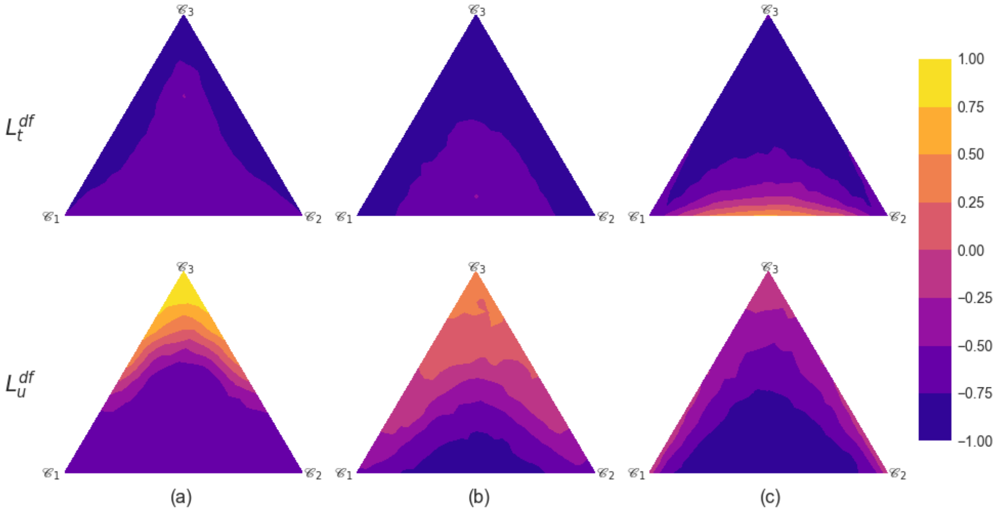


Figure 5.3: Effects of functional pressure alone across $\mathscr{C}$ for (a) $\lambda = 1$, (b) $\lambda = 10$ and (c) $\lambda = 20$. The upper-row, L_t^{df}, shows the mean difference between the highest proportion of target types and the highest other type in 1000 independent populations after 500 replicator steps for (a), (b), and (c). The lower-row, L_u^{df}, shows the mean difference between the highest proportion of competitor types and the highest other type in these populations.

between targets and competitors. For illustratory purposes, we focus on $b = .3$ in the following. Main differences in learnability instead come from the posterior parameter γ and the distribution over state distributions $\mathscr{C}$. As in Chapter 4 and in somewhat analogous fashion to the effects that λ has on signaling behavior, higher γ increases differences in the learnability of types. As for $\mathscr{C}$, the fidelity by which target types are transmitted is high relative to that of other types if at least two contexts are highly frequent. For example, if $\mathscr{C}_1 = .45 = \mathscr{C}_2$ target types are transmitted with a fidelity of approximately .6 for $\gamma = 1$ and .98 for $\gamma = 15$ ($k = 5, b = .3$). Their transmission fidelity is instead low if learners witness data predominantly in a single context. For example, if $\mathscr{C}_1 = .9$ and $\mathscr{C}_2 = \mathscr{C}_3$ then the probability of passing on target types diminishes to approximately .1 for $\gamma = 1$ and to almost zero if $\gamma = 15$ ($k = 5, b = .3$). This is expected given that learners that are frequently exposed to only c_1 or only c_2 have a hard time distinguishing whether their teachers' lexica are ambiguous. If $b > 0$, this lack of evidence increases the probability that learners acquire unambiguous lexica of the competitor kind. What is more, even without a bias for mutual exclusivity, unambiguous lexica are easier to learn because competitors' behavior is fairly deterministic compared to that of types with ambiguous lexica. Even if q is high, the latter will occasionally use different messages for the same state in the same context. As before, larger learning sequences (regulated by k) allow learners to

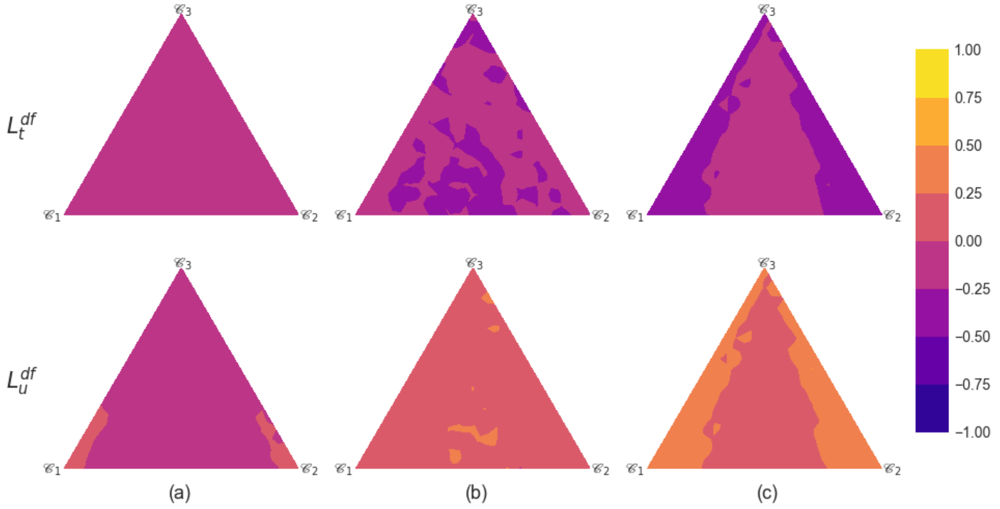


Figure 5.4: Effects of pressure for learnability alone across $\mathscr{C}$ with $\lambda = 20$ for (a) $\gamma = 1$ and $k = 5$, (b) $\gamma = 15$ and $k = 5$, and (c) $\gamma = 15$ and $k = 10$ ($b = .3$). The upper-row, L_t^{df}, shows the mean difference between the highest proportion of target types and the highest other type in 1000 independent populations after 500 mutator steps for (a), (b), and (c). The lower-row, L_u^{df}, shows the mean difference between the highest proportion of competitor types and the highest other type in these populations.

recover the type of their teacher with greater accuracy. The trends just mentioned nevertheless remain.

Figure 5.4 shows how pressure for learnability alone plays out. As above, this figure shows (i) the mean difference between the highest proportion of target types and the highest other type in 1000 independent populations, as well as (ii) the mean difference between the highest proportion of competitor types and the highest other type in these populations across values of $\mathscr{C}$ for two values of γ and k. As expected, target types do not fare well if there is no functional pressure at play. Competitor types fare better the higher γ, k, and – to a lesser degree – b are, but also fail to take over populations. As stressed in Chapter 4, pressure for learnability alone leads to the coexistence of multiple types, and consequently to highly polymorphic populations (see also Nowak 2006).

5.3.3 Functional pressure and learnability

We ascertained that neither pressure on its own leads to the prevalence of ambiguous target types. Nor does it lead to any other clear victor for that matter. While the expected utility of target types when communicating among themselves is high, functional pressure alone only leads to their selection in a small region

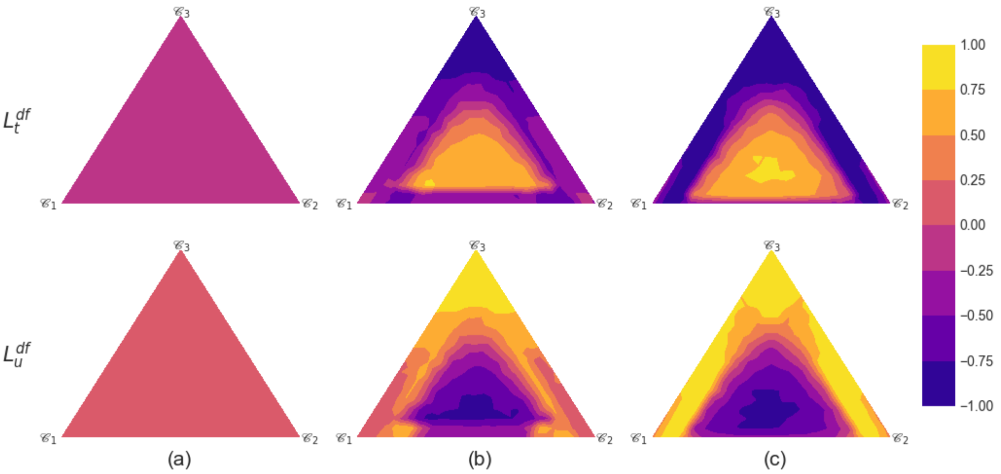


Figure 5.5: Effects of both pressures across $\mathscr{C}$ for (a) $\lambda = 1$, $\gamma = 1$ and $k = 5$, (b) $\lambda = 20$, $\gamma = 15$ and $k = 5$, and (c) $\lambda = 20$, $\gamma = 15$ and $k = 10$ ($b = .3$). The upper-row, L_t^{df}, shows the mean difference between the highest proportion of target types and the highest other type in 1000 independent populations after 500 replicator-mutator steps for (a), (b), and (c). The lower-row, L_u^{df}, shows the mean difference between the highest proportion of competitor types and the highest other type in these populations.

that favors contexts c_1 and c_2. Learnability disfavors target types and, as before, leads to polymorphic populations without a pronounced majority type. Next, we turn to the outcome effected by a joint pressure for learnability and for efficient and faithful information transfer.

Figure 5.5 shows the joint effect of both pressures for two values of λ, γ and k. As in Chapter 4 the emergence and stability of monomorphic populations is mainly influenced by λ and γ. If rationality is low and learners sample from the posterior (Figure 5.5a), there is not sufficient functional differentiation between types nor high transmission fidelity to allow for a stable and pronounced evolutionary outcome. By contrast, Figure 5.5b already showcases how the distribution over state distributions affects the evolutionary process: the center of the simplex favors near monomorphic populations of target types; its edges, and particularly frequent c_3, favor populations composed of unambiguous competitor types. This outcome is more pronounced for higher k because transmission fidelity increases (compare Figure 5.5b and Figure 5.5c).

These results stand to reason in light of our preceding discussion. Target types are easier to transmit in environments that allow them to showcase ambiguity exploitation of preferred m_3 to communicate s_1 in context c_1, in which this state is highly frequent; that of m_3 to communicate s_2 in context c_2 for analogous reasons; and ambiguity avoidance when in uniform c_3. The contribution of c_3 mainly lies in allowing learners to tease target types apart from more ambiguous types, while frequent communication in c_1 and c_2 confers them with a functional advantage over unambiguous competitors (see Figure 5.3). This advantage disappears if $\mathscr{C}_3$ is higher than $\mathscr{C}_1$ and $\mathscr{C}_2$.

In sum, the model predicts the emergence and stability of ambiguous target types, but not under all circumstances. Ambiguous target types evolve only if the world is such that agents find themselves in varied contexts; show a tendency to signal optimally according to their subjective perspective; and show a tendency to adopt the most likely type inferable from the overt behavior of their teachers.

5.4 General Discussion

Lexica that allow for the safe exploitation of preferred messages in informative contexts but lexicalize less ambiguous alternatives used to signal in uninformative ones can evolve and be taken up by a population. This can happen provided that the world is such that communication takes place in a mixture of these contexts. In general terms, this result reflects the fact that flexible types that can react to varied environments are typically favored over those that are narrowly specialized to few environments. Conversely, specialization wins over flexibility when there is little to no environmental variation. This is often true of biological as well as cultural evolution (Christiansen and Chater 2008:493).

These predictions add to the plausibility of our synchronic analysis of ambiguity in Chapter 3. The lexicon we assumed evolves in a mixture of the environments in which we predicted ambiguity exploitation to be functionally advantageous, and iterated communication to lead to coordination even without a common prior. Additionally, they strengthen the approach we followed in Chapter 4, where we argued that understanding phenomena at the semantics-pragmatics interface may require taking functional pressure as well as learnability into consideration. As in the case of scalar implicatures, either pressure on its own fell short from providing a justification for the pervasiveness of the property in question. However, the joint influence of both pressures suggests plausible conditions under which it emerges and stabilizes.

As for the particulars of ambiguity, learnability and in particular mutual exclusivity are important for they keep the transmission of more ambiguous lexica at bay. In this respect, learning plays a regularizing role. While more ambiguous lexica often lead to overt signaling behavior that is indistinguishable from that of target or competitor types, inferring them from the overt behavior of other agents is more difficult; even more so if there is at least a slight bias for mutual exclusivity. The contribution of functional pressure is straightforward and in line with what we have stressed throughout this and past chapters: it puts types in direct competition and promotes monomorphic populations. Under the right contextual conditions, this favors the selection of ambiguous target types and leads to their prevalence even if disfavored in learning.

The main difference between our application of the replicator-mutator dynamic in Chapter 4 and the present chapter concerns the involvement of a distribution over state distributions. Our motivation for its inclusion was to analyze the effects that the context of interaction has on language evolution where it is known to play an important role, as is the case for disambiguation. The connection between ambiguity and frequency is well established. Zipf (1949) already suggested that frequent words show a tendency to be associated with more meanings. Additionally, frequent words are typically short, predictable, and phonotactically unmarked (see, e.g., Dautriche 2015, Dautriche et al. 2017 and references within Chapter 3). Our results square well with this connection but do not support the idea that frequent words will inevitably be ambiguous. Instead, they suggest a qualification: ambiguity survives over time only when multiple meanings associated with a preferred form are frequent *and* appear in contrasting contexts in which one state is markedly more expected than the other. In our setup, state s_1 was frequent and expected in context c_1, and state s_2 was frequent and expected in context c_2. Frequent communication in both contexts is what (i) endows speakers of ambiguous lexica with a functional advantage over unambiguous ones and (ii) allows learners to infer that a message is semantically associated with two states from overt language use. The conclusion that message frequency unconditionally breeds ambiguity does not follow because the functional advantage of semantic ambiguity hinges on receivers being able to correctly infer different states across contexts. Either there are multiple contexts in which this is possible, or there is a single one in which a preferred message may lexicalize to signal the frequent meaning exclusively (see Figure 5.5 and Chapter 3). As shown in Figure 5.4, if a single context with a frequent state is more frequent, then an unambiguous lexicon is also easier to learn. According to our analysis, the relationship between frequency and ambiguity is consequently as follows: a preference for certain forms in language use leads to semantic ambiguity inasmuch as ambiguity is safe to be exploited in use and inferable by learners from their observable behavior. This leads frequent meanings to show a tendency to be associated with a single form if they appear in contrasting contexts, where these meanings tend to be recoverable.

The idea that the true distribution over states can have an impact on an evolving linguistic system has also been explored by Perfors and Navarro (2014), although only within the iterated learning tradition. That is, without a communicative task involving language use (see §4.3 for discussion). Perfors and Navarro's premise is nevertheless similar to ours: learning can be affected not only by the production and inference algorithms of teachers and learners, but also by the environment in which language is used. Differently from here, they not only assume that P^* affects the frequency in which data is produced but also that the observation of states (without accompanying linguistic material) is informative for the learner. In their own words, "it might be that language carries with it certain assumptions about what events are possible or probable in the world" therefore "simply observing meaningful events x may bias the learner to prefer some languages over others" (Perfors and Navarro 2014:779). In general terms, I agree with their assessment that the types in a population can be informative about the environment in which linguistic behavior takes place. After all, utility and fitness are functions of P^* and the population. In turn, they shape what types emerge, spread, and stabilize. In this sense, well-adapted types may show traces of the environment in which they emerged. However, the stronger hypothesis that the environment is informative *for learners* raises two intertwined issues.

First, focusing on learning only, it is an empirical question whether this information is used, or even extractable, by naïve learners who have yet to acquire a type. To the best of my knowledge, this claim has not been sufficiently addressed in the literature to decide one way or another. The good news is that the precision in which Perfors and Navarro's (2014) and our model are formulated allow for the investigation of this issue in a straightforward matter; by asking about the extent to which learners are aware of P^* prior to or during type acquisition; and, if they are aware at all, whether and how this affects learning.

Second, and more generally, it is doubtful that much can be learned from the environment alone for it to be informative about the types that interact in it. The reason is simply that it is hard, if not impossible, to read off which factor contributes to an evolutionary outcome and to what degree, by observing only the outcome itself. We have already seen that linguistic outcomes can result from non-trivial interactions between pressures that apply on cultural evolution under idealized conditions. Add more realistic complexity to these factors, as well as biological and social influence, and it seems difficult to maintain this hypothesis. As with the first issue, we do not have strong evidence to decide either way, but past and present research suggest caution on this front. We return to this issue in Chapter 6 for a broader assessment of linguistic outcomes, the processes likely to give rise to them, and what models can tell us about these matters.

Another analysis close in spirit is that of Santana (2014) who analyzed the evolution of ambiguity using the replicator(-mutator) dynamic as well. Differently from here, Santana stipulated a fixed mutation rate and assumed contextual information to always be informative about the state in play. That is, based on the contextual information at their disposition, receivers knew with certainty that some states did not obtain at a given interaction. We instead used (iterated) Bayesian learning to model transmission fidelity, assumed a common prior but no information about the particular state in play, and had varied objective state distributions that enabled for ambiguous signals to be exploited in multiple ways. This enabled us to tackle the challenge that the pervasiveness of ambiguity poses in light of the known problems it raises for language acquisition.

Our setup gives room for further analysis and refinement as we focused only on three concrete state distributions rather than on a larger space of distributions. Or, arguably even more naturally, on an infinite one. This choice was mainly driven by pragmatic considerations about simple setups. On the one hand, it is relatively straightforward to analyze the predictions of, for instance, Dirichlet distributed $\mathscr{C}$. On the other hand, the qualitative results reported above are not expected to be affected by this. However, it would certainly make the analysis and our exposition more complex. We have already discussed the choice to fix q and that of a common prior over states above, as well as that of common parameters such as λ, γ and k, together with possible refinements, in Section 4.5. At this point, I should reiterate that these choices are particular to the questions addressed. Different questions or additional empirical evidence might call for a relaxation or refinement of these assumptions (cf. Chapter 3).

5.5 Conclusion

This chapter addressed the question under which conditions (un)ambiguous lexical meanings (fail to) lexicalize. In particular, we focused on how variation in the context in which communication and learning take place can affect such evolutionary outcomes. Semantic ambiguity is predicted to evolve when the world is varied, enabling for the relatively safe pragmatic exploitation of preferred messages in contrasting informative contexts. A world that instead favors a single context promotes specialization over flexibility. Preferred messages are then predicted to be semantically associated with single states unambiguously, thereby reducing unnecessary risk introduced by uncertainty in signaling. The same is true of signaling among less rational agents, where specialization in the form of unambiguous lexical conventions safeguards against mistakes.

Chapter 6

General discussion

> Conventions are like fires: under favorable conditions, a sufficient concentration of heat spreads and perpetuates itself. The nature of the fire does not depend on the original source of heat. Matches may be the best fire starters, but that is no reason to think of fires started otherwise as any the less fires.
>
> David Lewis, *Convention: A Philosophical Study*

Communication is a complex social affair of which much is still little understood. It is therefore unsurprising that models of language, its use, and its transmission do not purport to provide fully accurate descriptions but involve substantial abstraction and simplification. We turn to such models not for their detail or exactness but for their explanatory force, with the goal "[...] to refine, systematize, and expand the menu of available explanations" (Ylikoski and Aydinonat 2014:23) by uncovering likely "systematic patterns of counterfactual dependence" (Woodward 2004:191). This investigation used models to better understand the relationship between factors that shape language (use) and properties or outcomes evidenced in natural language. In what follows, we reflect on our analysis along these more general lines. Section 6.1 discusses what the past chapters suggest about the relationship between semantics and pragmatics. Section 6.2 reviews methodological challenges faced by this kind of research. We argue that meeting these challenges calls for a pluralistic approach, in which we view our own efforts as being embedded.

6.1 On Semantics and Pragmatics

At the beginning of this investigation we posed two broad questions. The first question concerned the role of pragmatics in light of semantics. The second question reversely asked about the role of semantics in light of pragmatics. We can now reflect on how much headway we made in answering these questions.

Why leave to pragmatics what semantics can do? At first sight, it may seem as if semantic conventions ought to be as precise and constraining as possible. Under the assumption that such conventions are largely shared by interlocutors, this would leave less room for uncertainty and ensuing misunderstanding. The maxim for optimal linguistic design this view suggests is that to each meaning ought unequivocally correspond a unique form. Even under this view, it would still be useful to recruit mutual reasoning or contextual information in situations in which some uncertainty is unavoidable. Factors that may promote such supplementary reliance on pragmatics include open communities, changing and noisy environments, as well as change in semantic conventions effected by synchronic and diachronic processes yet to be adopted by all. In other words, even under the view that semantics should do much of the heavy lifting necessary for communication to succeed, pragmatics may still come into play to quench uncertainty where it is unavoidable.

This investigation focused on cases where relying on pragmatics was not necessary but an option; either in terms of linguistic choice or as a viable evolutionary outcome. In the case of ambiguity, we looked at lexica with unambiguous alternatives (Chapter 3), and analyzed the evolutionary competition of such lexica against less and more ambiguous alternatives (Chapter 5). In the case of scalar implicatures, we analyzed the evolutionary competition of lexica that rely on pragmatics to convey upper-bounds for weak scalar alternatives and those that enforce such bounds lexically (Chapter 4). The reason for looking at such cases was to elucidate whether pragmatics fulfills other roles as well. Roles where it comes into play not only due to environmental or biological constraints that necessitate it. Overall, the past chapters paint a different picture of the relationship between semantics and pragmatics than that of pragmatics playing solely a supplementary role.

Pragmatics, broadly construed, endows interlocutors with flexibility to react to different contexts of interaction by scaffolding on underspecified semantic conventions (Chapter 3 and 5). Intuitively, less precise semantics leave more room for pragmatic refinement. Such pragmatic refinement can offer better fits to the context of interaction and interlocutors involved in it than fixed precise semantics as it enables for linguistic material to be flexibly repurposed. Put differently, laxer semantic conventions endow pragmatic language users with the ability to convey a wide array of speaker meanings in an efficient, effective, and flexible manner. In this way, an utterance such as “I am cool” can inform the hearer about the speaker’s wellbeing; be used to decline an offer; describe temperature, popularity, trustworthiness; or something different altogether. Whether this ability confers its users with a functional advantage ultimately depends on the context(s) in which communication takes place (Chapter 3 and 5). Generally speaking, and ignoring side conditions such as agents’ rationality and their beliefs about each other’s expectations, the more varied yet informative contexts are, the more functionally advantageous it can be to rely on contextual information to guide inference.

Additionally, leaving to pragmatics what semantics could – in principle – do, may be explained not only in functional terms. In some cases, underspecified semantics may also be easier to acquire than more precise counterparts (Chapter 4). The explanation of pragmatic recruitment then lies not (solely) on communicative advantages that it can confer but on factors that shape the semantic conventions that enable for pragmatic inference in the first place. In this case, their learnability.

While in both Chapter 4 and Chapter 5 our analysis ultimately involved an interaction between functional pressure and learnability, there are some important differences between these explanations. In the first case we have an adaptive trait: the evolution of semantics that enable for pragmatic ambiguity exploitation is, to a large extent, a consequence of selection for greater communicative efficiency. If the environment allows for pragmatic inference to be advantageous then a balance between constraining semantic conventions and their pragmatic refinement is struck. The second case is not that of an adaptation per se, but rather a consequence of linguistic knowledge being shaped by its cultural transmission, with pragmatic reasoning enabling for the maintenance of underspecified semantics without incurring functional deficiencies.[1]

These predictions rest on relatively weak assumptions about agents' cognitive capacities. For ambiguity exploitation or scalar inferences to evolve, we only required agents to reason about each other's literal signaling behavior (level-1 reasoning); for them to exhibit a tendency toward utility maximization from their subjective perspective; and for them to exhibit a tendency toward posterior maximization in language acquisition. This is a desirable prediction when it comes to explanations of pervasive outcomes with multiple evolutionary starting points (Paternotte and Grose 2017). It suggests the (model internal) requirements for the emergence of these synergies between semantics and pragmatics to be relatively low, which is what we would expect of pervasive and cross-linguistically well attested properties.

Why leave to semantics what pragmatics can do? The above should not be taken to suggest that semantics plays a supplementary role to pragmatics either.

First, we should not forget that shared semantic conventions are often necessary to get pragmatic reasoning off the ground (Chapter 2). Silence, a grunt,

[1] One might argue that the learning algorithm employed in language acquisition or learners' inductive biases are adaptations on their own. This may well be so (see Christiansen and Chater 2008 for arguments to the contrary), but the point here is just that there is a difference. The former case, where there is a clear functional advantage, is an instance of "selection in action" (Ridley 2002:10): the type in question outcompetes alternatives in terms of communicative efficiency. The latter case is one where the type evolves by a combination of a functional advantage over some types and a learnability advantage over others that do not rely on pragmatic reasoning in communication.

or an utterance in a language the hearer does not speak arguably seldom constrain the space of possible meanings sufficiently to ensure that communication succeeds. Moreover, systematic pragmatic inferences, such as scalar implicatures, build on semantic constraints and their relationship to one another. As put by Searle (1971:190), often "meaning something when one says something is more than just contingently related to what the sentence means in the language one is speaking."

Second, our analysis suggests that there are conditions under which more constraining semantics emerge as natural and stable evolutionary outcomes. This can happen when the context of interaction is static, leaving no evolutionary grip for ad hoc pragmatic inferences to latch on underspecified semantics (Chapter 5); or it can happen when rationality is low (Chapter 4 and 5). In either case semantic specialization wins over the potential for refinement that pragmatics fosters.

In the case of a static context, e.g., when there is a single context governed by a fixed distribution over states, populations may adopt functionally efficient semantic conventions that do not rely on pragmatics but are semantically well-tailored to the context. In other words, if there are no situations in which the flexibility that pragmatics enables for can be cashed out then leaving to pragmatics what (shared) semantics can do is either functionally disadvantageous; or, at best, a neutral trait (Chapter 3 and 5). Moreover, in this case semantics that are narrowly specialized to the context are also easier to acquire than those that harbor ambiguity (Chapter 5).

The other important case in which semantic precision may be favored over pragmatic recruitment is relative to agents' rationality. Pragmatic inference involving mutual reasoning needs to be fueled by some degree of rationality. Less constraints on the semantic side ask more of reasoners: they not only need to learn the semantic conventions of their community but also need to appropriately deploy them pragmatically to achieve communicative goals. Less rational agents are accordingly better served with more constrained form-meaning associations that leave less room for uncertainty and misunderstanding. Reversely, less is often more if agents are rational (in tendency, see Chapter 4 and 5 for qualification).

Taking stock, we can see semantics as providing (near-)global constraints on form-meaning associations taken to be shared by a community. Pragmatics endows language users with the ability to refine upon them locally, depending on the context of interaction and on whom one interacts with (Chapter 3). Of course, for pragmatic language use to lead to successful coordination, the mechanisms through which semantic conventions are refined also need to be shared and mutually recognized. If the sophistication of agents is low or the environment is static enough for the distinction between global and local to collapse then semantics can take over most of the functions suggested above. Reversely, communication of ecologically rational agents in rich environments foments the kind of divisions of labor between semantics and pragmatics we see in natural language.

While these results may be intuitive after the fact, we showed that divisions of labor between semantics and pragmatics arise from complex interactions between (i) agents' cognitive make-ups (signaling behavior or learning mechanisms), (ii) relevant pressures (functional pressure on efficient communication or learnability), and (iii) the context(s) in which communication takes place. The outcome of such interactions can effect linguistic change that is adopted by a population, but also to the coexistence of different divisions of labor (Chapter 4). The latter result is important as it highlights the fact that semantic conventions and pragmatic rules operating on them are unobservable by themselves. Language users and learners witness only the observable effects of their interaction. Different divisions of labor can therefore lead to largely indistinguishable overt signaling behavior and coexist.

6.2 Change, Outcomes, and Factors of Influence

Just as biological evolution, linguistic change has no foresight. If agents adapt and thereby optimize information transfer, they do so with respect to their interlocutors and the context of interaction in the present (Chapter 3; Pate and Goldwater 2015); not with regard to longer term optimizations of their language (use). That is to say that pervasive properties of natural language (use) are not products of explicit deliberation or design undertaken by their users. As already noted in the introduction, the behavior of an individual at a given time is consequently not necessarily informative about the effects that linguistic pressures have (had) on her language and behavior in longer time stretches.

In Section 2.2 we expanded on this issue from a methodological perspective. Among others, we argued that explanations of linguistic properties need be explicit about how relevant factors that may give rise to them interact. First, this is necessary to add force to the explanation of a linguistic property. An explanatory analysis should add to our understanding of the conditions under which a property comes to light or changes, and why this happens. Second, modeling the interaction of relevant factors explicitly is necessary because they can interact in non-trivial ways, rendering direct inference from factors to outcomes difficult, if not impossible. Framed more positively: a better understanding of linguistic properties is gained through the inspection of the interaction of factors such as individual-level behavior, the environment in which communication takes place, the communicative task at hand, population dynamics and transmission perturbations that affect how linguistic knowledge is passed on.

Analogous difficulties are faced in the opposite direction: if the interaction between language (use) and pressures that apply on it are non-trivial, then we are seldom justified to draw strong inferences from outcomes about the factors that may have caused them.

While well known, we raise these general issues because they easily creep into evolutionary analysis. More so if it is seemingly intuitive. One illustrative case where there is growing interest in using evolutionary outcomes as diagnostics for underlying causes is found in the iterated learning literature. Recall that even weak inductive learning biases can have striking effects on an evolving linguistic system (Chapter 4 and 5). Much effort has accordingly been devoted to investigating what kind of biases there are. Some prominent examples are mutual exclusivity (Merriman et al. 1989, Clark 2009; see Chapter 5), simplicity (Chater and Vitányi 2003, Kirby et al. 2015; see Chapter 4), regularization (Hudson Kam and Newport 2005) and generalization (Smith 2011). If inductive biases can come to play such an important role, we might reasonably expect outcomes of iterated learning to show traces of the biases that shaped them. That is, we might expect "that systems of knowledge or behaviour [such as language and its properties] transmitted by iterated learning evolve to reflect the biases of individuals involved in transmission" (Kirby et al. 2014:110; see also, e.g., Kalish et al. 2007). As already argued in Chapter 5 in relation to Perfors and Navarro 2014 and rephrased above, it is nevertheless questionable how much we can learn about biases, or any other factors, *from an outcome alone*. Iterated learning, one may want to argue, presents a special case in this respect. The mathematical characterization of Griffiths and Kalish (2005; 2007) suggests that, under certain conditions, language evolution through iterated learning converges to the prior.[2]

Accordingly, we may expect human experiments using iterated learning to not only show traces of but actually reveal which biases are at play (e.g., Jacoby and McDermott 2017). This idea is problematic without qualification. Learning and typology certainly influence each other. After all, learnability is a necessary condition for culturally transmitted properties to see the light of day. However, learning outcomes and typology are by no means faithful reflections of each other (Bowerman 2010), and, more often than not, "iterated learning is doing more than just revealing the prior biases of learners" (Cornish 2011:173), as discussed in Section 4.5. Additionally, other factors and forces than learning biases may also systematically perturb the transmission of linguistic knowledge and thereby contribute to the shaping of language. Beyond the role that functional pressure on efficient communication plays, we saw that state frequencies can also affect the iterated transmission of linguistic knowledge (Chapter 5; see also Perfors and Navarro 2014). Additionally, whether learners tend to maximize the posterior (Kirby et al. 2007), the size of the population from which learners receive their input as well as whether biases are heterogeneous (Ferdinand and Zuidema 2009) can, among many others, also influence whether outcomes come to reflect inductive biases faithfully.

[2] In a nutshell, the main conditions are that there is a transmission chain – not influenced by factors beyond learning itself, such as language use or the environment –, that every agent uses the same production and learning algorithms, that they all have the same prior, and that they sample from the posterior ($\gamma = 1$ in our notation; see Griffiths and Kalish 2005; 2007 for details).

Another source of transmission perturbation that we have neglected so far but deserves brief mention in relation to this discussion is noisy perception: agents' imperfect perception of the world. The general idea is straightforward. If the world is not always perceived accurately, regular stochastic errors in the perception of states can lead teachers to produce utterances that deviate from their production behavior had they witnessed the state correctly. Similarly, learners may mistake utterances as applying to different states than the ones witnessed by the teacher who produced them. For instance, when learning the meaning of a vague adjective such as *tall* from utterances like "Jean is tall", agents may have diverging representations of how tall Jean actually is, even if she is in a shared perceptual environment. Over time, this may lead to the emergence of certain linguistic properties, in this case vagueness, not for functional reasons nor because of an inductive bias, but solely due to perceptual factors.

In Brochhagen and Franke (2017), we looked at some effects that noisy perception can have on iterated learning. The three simple case studies we analyzed are reproduced in Appendix C. The finding relevant to this discussion is that, indeed, regularities in misperceptions of states can have striking and possibly explanatory effects on language evolution. Such misperceptions can lead to biases of inferring the "wrong" teacher type if noise makes some types err in a way that resembles the noiseless behavior of other types. That is, such an environmental factor can, in principle, induce transmission perturbations that look as if there was a cognitive bias in favor of a particular type, simply because that type better explains the noisy behavior.

We mention noisy perception to underscore the issue raised by the relation between linguistic outcomes and their causes. If our arguments in the past chapters and the results that followed from them are on the right track, then actual communication and the environment in which it takes place can play non-negligible roles in shaping a linguistic outcome. Reasoning on the basis of a subset of factors will deliver correct explanations only if we accurately identified the ones that are relevant to the phenomenon at hand. Taken together with the difficulty of disentangling the effects that different factors have on a complex evolving system such as language, our general methodological takeaway is that a pluralistic stance needs to be adopted. What is needed are models that allow us to ask whether linguistic phenomena are due to learning (biases), environmental factors, functional pressure, or interactions thereof. The models we proposed in this investigation allow us to do exactly this. The application of the replicator-mutator dynamic in Chapter 4 can tease apart functional pressure and effects of iterated learning; its variant in Chapter 5 allows for the analysis of the effects that different state frequencies have on communication and learning; and the noisy iterated Bayesian learning model in Appendix C is a neutral model of cultural evolution that appeals to neither functional competition nor differential learnability among types (see also Reali and Griffiths 2009). As showcased throughout this investigation, this family of models is well compatible with probabilistic models of language use at the individual level, or other formalisms of linguistic choice for that matter. This speaks to their applicability to a wide-range of questions concerning natural language, its use, and its transmission. This investigation applied them to novel questions about the relationship between semantics and pragmatics, delivering some, if modest, answers.

Chapter 7

Conclusion

> Human rational behavior (and the rational behavior of all physical symbol systems) is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor.
>
> Herbert A. Simon, *Invariants of Human Behavior*

This investigation focused on the effects that solutions to two fundamental problems have on a changing linguistic system. The first problem is that of successful and efficient information transfer between boundedly rational agents. The second problem is that of the transmission of linguistic knowledge from proficient agents to naïve ones, who have not yet come to acquire semantic conventions or pragmatic dispositions to act on them. Solutions to both problems influence each other. The knowledge acquired by learners is, to a large extent, a product of solutions to the first problem. In turn, the linguistic means available to agents to solve the first problem are influenced by what they learn from others. We focused on the bearing of these solutions on conditions under which reliance on pragmatic inference is (not) favored over less equivocal semantic codification of information to be conveyed. In a nutshell, we showed that pragmatic recruitment in tandem with semantic underspecification offers greater flexibility to repurpose linguistic material, and that it can allow for the maintenance of simpler semantics that are easier to learn. In both cases communication comes to leverage contextual information and mutual reasoning – the unsaid – even if matters could be communicated more explicitly. Reversely, low rationality in choice or little variation in the information context provides can lead to stronger reliance on semantics over pragmatics.

In Chapter 3, we analyzed under which conditions safe functional exploitation of semantic ambiguity can come about. In particular, we showed that even when contextual expectations vary from one agent to the next, ambiguity can be a useful property for semantic conventions to harbor; because it allows interlocutors to flexibly repurpose linguistic material to better suit their communicative preferences. Our analysis predicts that the challenge faced when not having a common contextual prior is overcome if interlocutors interact multiple times, thereby enabling them to adapt their linguistic behavior to each other. These predictions are, by and large, borne out in experimental data.

A particularly interesting issue left open in this chapter concerns the dialog-initial adoption of Anti-Horn-like signaling, as attested in Kanwal et al.'s (2017) data. Neither the model I proposed nor the experimental setup of Kanwal et al. explain this behavior satisfactorily. More importantly, I think that there is more to be said about it than that it is a mere product of chance or error. As stressed in Section 3.5, to understand why speakers start an interaction by associating an infrequent meaning with a preferred form, a better understanding of the formation of contextual expectations and beliefs about them is needed.

Chapters 4 and 5 looked at population-level dynamics at the interface between semantics and pragmatics. Chapter 4 focused on factors that may explain a division of labor between them. In particular, on scalar implicatures and their (lack of) lexicalization. Chapter 5 asked under which contextual conditions of use and learning underspecified semantic conventions are both learnable and functionally efficient, thereby coming to ingrain themselves in a population. To answer these questions, we proposed an application of the replicator-mutator dynamic, which tracks effects of functional pressure on successful and efficient communication and pressure for learnability. We showed how this model can be combined with probabilistic models of rational language use and highlighted the kind of novel questions about change at the semantics-pragmatics interface it allows us to ask. In particular, by taking the challenge serious that neither semantics nor pragmatics are directly unobservable; only the behavior effected by their joint interaction is available for agents to base their inference on.

One of the main open issues from Chapter 4 concerns the inductive learning bias in favor of simpler lexical representations we assumed. The good news is that much effort is being devoted to the development of diagnostics and studies to further our understanding of the nature and relationship between lexical representations, their complexity, and their acquisition. For the time being, this assumption should however be seen critically. This will hopefully motivate future research to speak to this matter.

Finally, I would like to see the analysis in Chapter 5 be supplemented with actual frequency data. This would add strength to the prediction that (lexical) ambiguity is diachronically persistent if meanings attached to a form are frequent and tend appear in contrasting contexts.

While, hopefully, progress was made, these open issues show that this investigation is not exhaustive. In terms of particular phenomena, I have answered some questions; and hope that the material worked out in each chapter goes some way to address more; or that it at least provides a good starting point to ask fruitful ones.

In more general terms, when it comes to the semantics-pragmatics interface what this investigation offered is an outline of future directions to take and a number of tools to embark in them. On this front, our contribution lies in the modular characterization that allows for isolated and joint inspection of functional pressure and learnability on evolutionary trajectories, as well for the inspection of how frequency and perception modulate these forces and, together, shape linguistic knowledge and behavior. For now only a rather constrained space of phenomena and factors was explored, leaving us with the modest predictions formulated in Chapter 6.

Appendix A

The following tables supplement the simulation results reported in Section 3.4. All outcomes correspond to the mean of 10^4 independent simulations with $\lambda = 20$, $c_\sigma(m_1) = .4 = c_\sigma(m_2)$, and $c_\sigma(m_3) = .1$. The results in Table A.1 were obtained from the simple update mechanism of $\mathscr{P}$ presented in Section 3.3, in (3.5), and those in Table A.2 from its "preemptive" refinement in Section 3.4, in (3.9).

		10 iterations		30 iterations		50 iterations		
r	$P^*(s_1)$	EU_σ (SD)	JSD	EU_σ (SD)	JSD	EU_σ (SD)	JSD	$\text{EU}_\sigma^{\max}$
0.1	0.5	.61 (.08)	.03	.62 (.07)	.00	.62 (.06)	.00	.75
	0.7	.62 (.09)	.03	.64 (.09)	.01	.67 (.10)	.00	.81
	0.9	.64 (.10)	.03	.68 (.12)	.01	.71 (.13)	.00	.87
0.5	0.5	.59 (.08)	.01	.61 (.06)	.01	.61 (.06)	.00	.75
	0.7	.62 (.10)	.01	.66 (.10)	.00	.68 (.10)	.00	.81
	0.9	.66 (.12)	.00	.70 (.13)	.00	.72 (.13)	.00	.87
1	0.5	.59 (.08)	.02	.61 (.06)	.01	.61 (.06)	.00	.75
	0.7	.62 (.10)	.01	.66 (.10)	.00	.68 (.10)	.00	.81
	0.9	.67 (.12)	.01	.70 (.13)	.00	.72 (.13)	.00	.87

Table A.1: Mean sender expected utility and JSD of interlocutors' priors in 10^4 independent games. $\text{EU}^{\max}$ indicates the maximum expected utility reachable for a given P^*.

		10 iterations		30 iterations		50 iterations		
r	$P^*(s_1)$	EU_σ (SD)	JSD	EU_σ (SD)	JSD	EU_σ (SD)	JSD	$\text{EU}_\sigma^{\max}$
0.1	0.5	.59 (.12)	.03	.59 (.12)	.01	.58 (.11)	.02	.75
	0.7	.67 (.14)	.03	.76 (.09)	.01	.79 (.05)	.00	.81
	0.9	.77 (.12)	.03	.86 (.01)	.01	.87 (.00)	.00	.87
0.5	0.5	.58 (.12)	.02	.58 (.12)	.03	.58 (.11)	.03	.75
	0.7	.72 (.11)	.01	.79 (.05)	.00	.80 (.02)	.00	.81
	0.9	.83 (.05)	.01	.87 (.00)	.00	.87 (.00)	.00	.87
1	0.5	.58 (.12)	.02	.58 (.12)	.03	.58 (.12)	.03	.75
	0.7	.73 (.11)	.01	.79 (.05)	.00	.80 (.02)	.00	.81
	0.9	.84 (.04)	.00	.87 (.00)	.00	.87 (.00)	.00	.87

Table A.2: Mean sender expected utility and JSD of interlocutors' priors in 10^4 independent games using "preemptive" belief updates. $\text{EU}^{\max}$ indicates the maximum expected utility reachable for a given P^*.

Appendix B

95% highest posterior density and mean of subjects' marginal posterior for λ together with predicted RMSE. Each pair of subjects corresponds to a dyad in the game.

	Subject 1		(RMSE = 0.197)		**Subject 2**		(RMSE = 0.188)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	22.479	31.040	4.499	39.788	23.254	31.951	4.783	41.873

	Subject 3		(RMSE = 0.196)		**Subject 4**		(RMSE = 0.183)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	14.181	19.866	2.901	25.333	22.650	32.095	4.951	41.828

	Subject 5		(RMSE = 0.009)		**Subject 6**		(RMSE = 0.136)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	25.421	35.808	5.365	46.263	23.728	33.872	5.072	43.541

	Subject 7		(RMSE = 0.007)		**Subject 8**		(RMSE = 0.348)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	25.673	35.844	5.402	46.491	5.162	7.573	1.307	10.193

	Subject 9		(RMSE = 0.162)		**Subject 10**		(RMSE = 0.011)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	17.310	29.285	5.680	39.157	25.546	35.978	5.392	46.306

	Subject 11		(RMSE = 0.155)		**Subject 12**		(RMSE = 0.161)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	16.759	24.628	4.166	32.830	21.273	29.875	4.695	39.553

	Subject 13		(RMSE = 0.145)		**Subject 14**		(RMSE = 0.418)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max

λ	22.626	30.994	4.502	39.993	8.348	12.018	1.857	15.461

	Subject 15		(RMSE = 0.493)		**Subject 16**		(RMSE = 0.433)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	8.708	12.435	2.041	16.508	12.257	17.100	2.670	22.556

	Subject 17		(RMSE = 0.308)		**Subject 18**		(RMSE = 0.394)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	20.218	28.133	4.172	36.612	9.110	13.258	2.095	17.424

	Subject 19		(RMSE = 0.208)		**Subject 20**		(RMSE = 0.0)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	12.402	16.867	2.425	21.751	21.042	31.133	5.217	41.414

	Subject 21		(RMSE = 0.191)		**Subject 22**		(RMSE = 0.189)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	16.468	23.265	3.490	30.184	17.145	23.318	3.438	30.577

	Subject 23		(RMSE = 0.189)		**Subject 24**		(RMSE = 0.191)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	16.054	23.036	3.634	30.077	16.788	23.296	3.559	30.529

	Subject 25		(RMSE = 0.188)		**Subject 26**		(RMSE = 0.484)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	15.304	22.383	3.727	29.752	3.738	5.314	0.796	6.826

	Subject 27		(RMSE = 0.326)		**Subject 28**		(RMSE = 0.191)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	19.914	27.638	4.316	36.331	16.659	23.159	3.627	30.655

	Subject 29		(RMSE = 0.19)		**Subject 30**		(RMSE = 0.188)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	15.358	22.500	3.738	29.411	16.520	22.660	3.586	30.160

	Subject 31		(RMSE = 0.189)		**Subject 32**		(RMSE = 0.188)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	15.013	22.051	3.707	29.164	15.343	22.348	3.706	29.927

	Subject 33		(RMSE = 0.15)		**Subject 34**		(RMSE = 0.453)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	17.223	24.629	3.849	31.969	4.259	6.055	0.990	8.090

	Subject 35		(RMSE = 0.322)		**Subject 36**		(RMSE = 0.414)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	6.998	10.183	1.783	14.000	6.605	9.444	1.614	12.743

	Subject 37		(RMSE = 0.0)		**Subject 38**		(RMSE = 0.0)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	21.435	31.147	5.074	40.896	21.266	31.212	5.169	41.269

	Subject 39		(RMSE = 0.227)		**Subject 40**		(RMSE = 0.141)	
	HPD min	mean	SD	HPD max	HPD min	mean	SD	HPD max
λ	15.153	21.749	3.429	28.416	25.497	35.532	5.307	45.470

Appendix C

This appendix reproduces the noisy iterated learning model from Brochhagen and Franke 2017 together with the three illustrative case studies found therein.

Iterated Bayesian Learning with State-Noise

Other stochastic factors beyond learning biases in $P(\tau)$ can influence the adoption of a linguistic type τ based on the observation of $\langle s, m\rangle$-sequences. One further potential source of "transmission noise" are regular stochastic errors in the perception of world states.

We denote the probability that the teacher (learner) observes state s_t (s_l) when the actual state is s_a as $P_N(s_t \mid s_a)$ ($P_N(s_l \mid s_a)$). The probability that s_a is the actual state when the learner observes s_l is therefore:

$$P_N(s_a \mid s_l) \propto P(s_a)\ P_N(s_l \mid s_a)\,.$$

Assuming a finite state space for convenience, the probability that the teacher observes s_t when the learner observes s_l is:

$$P_N(s_t \mid s_l) = \sum_{s_a} P_N(s_a \mid s_l)\ P_N(s_t \mid s_a)\,.$$

The probability that a teacher of type τ produces data that is perceived by the learner as a sequence d_l of $\langle s_l, m\rangle$-pairs is:

$$P_N(d_l \mid \tau) = \prod_{\langle s_l, m\rangle \in d_l} \sum_{s_t} P_N(s_t \mid s_l)\ P(m \mid s_t, \tau)\,.$$

We assume that learners, even if they (in tendency) perform rational Bayesian inference of the likely teacher type τ based on observation $\langle s_l, m\rangle$, do not also reason about state-noise perturbations. In contrast to, e.g., noisy-channel models that have agents reason over potential message corruption caused by noise (e.g. Bergen and Goodman 2015), our learners are not proficient language users that could leverage knowledge about the world and its linguistic codification to infer likely state misperception.[1] In this case the posterior probability of τ given the learner's perceived data sequence d_l is as before: $P(\tau \mid d_l) \propto P(\tau)\ P(d_l \mid \tau)$. Still, state-noise affects the probability $P_N(\tau_j \to \tau_i)$ that the learner adopts τ_i given a teacher of type τ_j, because it influences the probability of observing a sequence d_l (with $F(\tau_i \mid d)$ as before):

$$P_N(\tau_j \to \tau_i) \propto \sum_{d \in D_k} P_N(d_l \mid \tau_j) F(\tau_i \mid d)\,.$$

Noise free iterated Bayesian learning is obtained as a special case when the perceived state is always the actual state.

Case Studies

We present three case studies that show how iterated learning under noisy perception can lead to the emergence of linguistic phenomena. The studies are ordered from more to less obvious examples in which state-noise may be influential and explanatory: (i) vagueness, (ii) meaning deflation, and (iii) underspecification in the lexicon.

No case study is meant to suggest that state-noise is the definite and only explanation of the phenomenon in question. Instead, our aim is to elucidate the role that transmission perturbations beyond inductive biases may play in shaping the cultural evolution of language. We therefore present minimal settings that isolate potential effects of state-noise in iterated learning.

Vagueness

Many natural language expressions are notoriously vague and pose a challenge to logical analysis of meaning (e.g., Williamson 1994). Vagueness also challenges models of language evolution since functional pressure toward maximal information transfer should, under fairly general conditions, weed out vagueness (Lipman 2009). Many have therefore argued that vagueness is intrinsically useful for communication (e.g., van Deemter 2009, De Jaegher and van Rooij 2011, Franke et al. 2011, Blume and Board 2014). Others hold that vagueness arises naturally due to limits of perception, memory, or information processing (e.g., Franke et al. 2011, O'Connor 2014, Lassiter and Goodman 2015). We follow the latter line of exploration here, showing that vagueness can naturally arise under imperfect observability of states (see Franke and Correia 2017 for a different evolutionary dynamic based on the same idea).

[1] To do so, agents would have to infer or come equipped with knowledge about $P_N(\cdot|s_a)$, which could itself be subject to updates. We stick to the simpler case of noise-free inference here, but as long as the actual state is not always recoverable our general results also hold for agents that reason about noise.

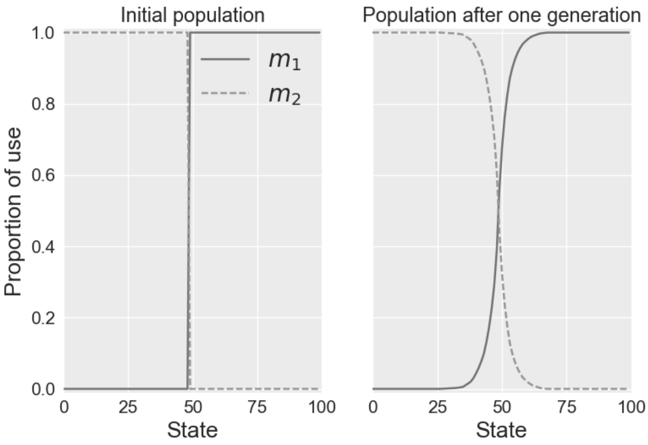


Figure C.1: Noisy iterated learning ($\gamma = 1$, SD $= 0.4$, $k = 20$).

Setup. We analyze the effects of noisy perception on the transmission of a simple language with 100 states, $s \in [0; 99]$, and two messages, $m \in \{m_1, m_2\}$. The probability that agents perceive actual state s_a as s_t/s_l is given by a (discretized) normal distribution, truncated to $[0; 99]$, with s_a as mean and standard deviation SD. Linguistic behavior is fixed by a type $\tau \in [0; 99]$ which is the threshold of applicability of m_1: $P(m_1 \mid s, \tau) = \delta_{s \geq \tau} = (1 - P(m_2 \mid s, \tau))$. In words, if a speaker observes a state that is as large or larger than its type, then message m_1 is used (e.g., *tall*), otherwise m_2 is used (e.g., *small*).

Results. The effects of a single generational turnover under noisy transmission of a population that initially consisted exclusively of type $\tau = 50$ is depicted in Figure C.1. As learners try to infer this type from observed language use, even small SD will lead to the emergence of vagueness in the sense that there is no longer a crisp and determinate cut-off point for message use in the population. Instead, borderline regions in which m_1 and m_2 are used almost interchangeably emerge. For larger SD, larger borderline regions ensue. The size of such regions further increases over generations with growth inversely related to γ and k. As is to be expected, if k is too small for learners to discern even strikingly different types, then iterated learning under noisy perception leads to heterogeneous populations with (almost) no state being (almost) exclusively associated with m_1 or m_2.

Discussion. Transmission perturbations caused by noisy state perception reliably give rise to vague language use even if the initial population had a perfectly crisp and uniform convention. Clearly, this is a specific picture of vagueness. As modeled here for simplicity, each speaker has a fixed and non-vague cut-off point τ in her lexicon. Still, the production behavior of a type-τ speaker in actual state

s_a is probabilistic and "vague", because of noisy perception:

$$P_N(m \mid s_a, \tau) = \sum_{s_t} P(s_t \mid s_a) P(m \mid s_t, \tau)\,.$$

An extension toward types as distributions over thresholds is straightforward but the main point would remain: systematic state-noise perturbs a population toward vagueness.

Of course, convergence on any particular population state will also depend on the functional (dis)advantages of particular patterns of language use. Functional pressure may therefore well be necessary for borderline regions to be kept in check, so to speak. Which factor or combination thereof plays a more central role for the emergence of vagueness is an empirical question we do not address here. Instead, we see these results as adding strength to the argument that one way in which vagueness may arise is as a byproduct of interactions between agents that may occasionally err in their perception of the environment. If state perception is systematically noisy and learners are not aware of this, some amount of vagueness may be the natural result.

Deflation

Meaning deflation is a diachronic process by which a form's once restricted range of applicability broadens. Perhaps the most prominent example is Jespersen's cycle (Dahl 1979), the process by which emphatic negation, such as French *ne ... pas*, broadens over time and becomes a marker for standard negation. As argued by Bolinger (1981), certain word classes are particularly prone to slight and unnoticed reinterpretation. When retrieving their meaning from contextual cues, learners may consequently continuously spread their meaning out. For instance, Bolinger discusses how the indefinite quantifier *several* has progressively shifted from meaning *a respectable number* to broader *a few* in American English. We follow this line of reasoning and show how state confusability may lead to meaning deflation. Other formal models of deflationary processes in language change have rather stressed the role of conflicting interests between interlocutors (Ahern and Clark 2014) or asymmetries in production frequencies during learning (Schaden 2012, Deo 2015).

Setup. The setup is the same as that of the previous case study, except that we now trace the change of a single message m, e.g., emphatic negation, without a fixed antonym being sent whenever m does not apply. This is a crude way of modeling use of markers of emphasis or high relevance for which no corresponding "irrelevance marker" exists. Learners accordingly observe positive examples of use $\langle s, m \rangle$ but do not positively observe situations in which m did not apply to a particular state. This causes asymmetry in the learning data because some types will reserve their message only for a small subset of the state space and otherwise remain silent. Learners take the absence of observations into account but cannot know what it is that they did not observe. We assume that learners are aware of k so that:[2]

$$P(\tau|d_l) \propto \text{Binom}(\text{successes} = k - |d_l|, \text{trials} = k, \\ \text{succ.prob} = \sum_{i=0}^{\tau-1} P(s = i)) \prod_{s \in d_l} P(m|s, \tau)\,.$$

As before, the second factor corresponds to the likelihood of a type producing the perceived data. The first is the probability of a type not reporting $k - |d|$ events for a total of k events. $P \in \Delta(S)$ is assumed to be uniform. In words, a long sequence of data consisting of mostly silence gives stronger evidence for the type producing it having a high threshold of applicability even if the few state-message pairs observed may be equally likely to be produced by types with lower thresholds.

Results. The development of an initially monomorphic population consisting only of $\tau = 80$ is shown in Figure C.2. Even little noise causes a message to gradually be applied to larger portions of the state space. The speed of meaning deflation is regulated by SD, k, and to lesser degree γ. In general, more state confusion due to higher SD, shorter sequences, or less posterior maximization will lead to more learners inferring lower types than present in the previous generation.

Discussion. In contrast to the previous case study, we now considered the effects of noisy perception under asymmetric data generation where overt linguistic evidence is not always produced, i.e., acquisition in a world in which not every state is equally likely to lead to an observable utterance. The outcome is nevertheless similar to the previous one: Noisy perception can cause transmission perturbations that gradually relax formerly strict linguistic conventions. In contrast to the case of vagueness, if there are no relevant competing forms, e.g., *small* vs. *tall*, asymmetry in production and noise will iteratively increase the state space that a form carves out.

Scalar Expressions

Why does regular pragmatic strengthening not lead to wide-spread lexicalization of upper-bounded meanings in weak scalar expressions? To address this question,

[2] Knowing k allows learners to compute the likelihood of a type not reporting $k - |d_l|$ state observations. A better but more complex alternative is to specify a prior over k with learners performing a joint inference on k and the teacher's type. For simplicity, we opt for the former, albeit admittedly artificial, assumption.

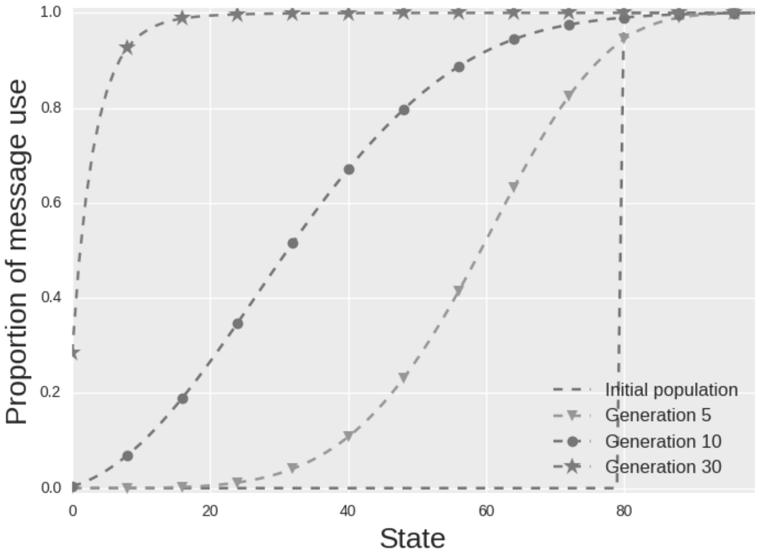


Figure C.2: Noisy iterated learning ($\gamma = 1$, SD $= 0.4$, $k = 30$).

in Chapter 4 we explored an evolutionary model that combines functional pressure and iterated learning. This analysis assumed a prior that favors a lack of upper-bounds relative to their lexicalization. Here, we demonstrate that state-noise can mimic the effects of such a cognitive learning bias in a reduced type space consisting only of users of lexica L_{bound} and L_{lack}. As in Chapter 4, a type is a pair of either of these two lexica and level-0 or level-1 behavior (see Chapter 4 for details).

Setup. As pragmatic users of L_{lack} are (almost) indistinguishable from types with L_{bound}, the emergence of a predominance of L_{lack} in a repeatedly learning population must come from transmission biases. A learning bias in favor of L_{lack} in the learners' priors will lead to its predominance (Chapter 4), but here we assume no such cognitive bias. Rather we assume state-noise in the form of parameters ϵ and δ. The former is the probability of perceiving actual state $s_{\exists\neg\forall}$ as $s_{\forall}$, $P(s_{\forall}|s_{\exists\neg\forall}) = \epsilon$, and $P(s_{\exists\neg\forall}|s_{\forall}) = \delta$. For instance, states may be perceived differently because different numbers of objects must be perceived (e.g., quantifiers and numerals) or they may be more or less hard to accurately retrieve from sensory information (e.g., adjectives).

Results. To quantify the effects of the dynamics we ran a fine-grained parameter sweep over ϵ and δ with 50 independent simulations per parameter configuration. Each simulation started with a random initial population distribution over types and applied iterated learning with state-noise for 20 generations, after which no noteworthy change was registered. Mean proportions of resulting

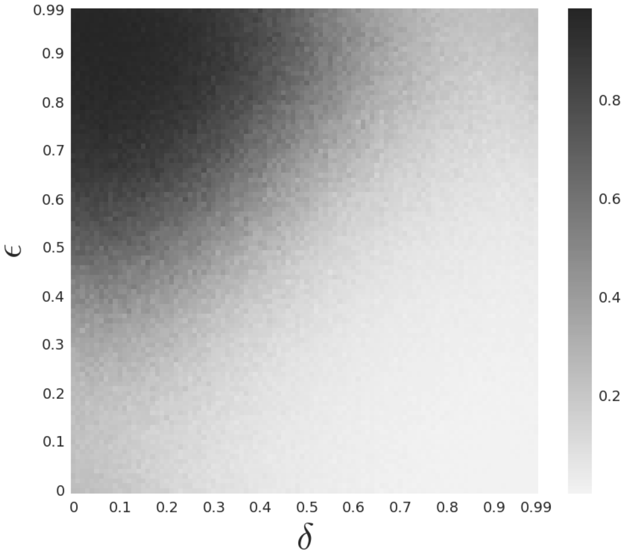


Figure C.3: Mean proportion of pragmatic L_{lack} users after 20 generations ($\gamma = 1$, $k = 5$).

pragmatic users of L_{lack} under different noise signatures are shown in Figure C.3. These results suggest that when δ is small and ϵ high, iterated noisy transmission can lead to populations consisting of almost exclusively English-like lexica with pragmatic language use. Similar results are obtained for larger k or γ.

Discussion. The main goal of this case study was to show that noisy perception may mimic effects of learning biases. In the case of Chapter 4 the assumed bias was one for simplicity; learners had an a priori preference for not codifying upper-bounds lexically, which increased their propensity to infer pragmatic L_{lack} over L_{bound} even if the witnessed data could not tease them apart. We assumed no such bias but nevertheless arrived at evolutionary outcomes that are comparable to those predicted if the bias were present. However, this result strongly depends on the types involved. Whether a type thrives under a particular noise signature depends on the proportion of types confused with it during transmission. The addition or extraction of a single type therefore leads to different results.

I should stress that the evolution of weak scalar expressions lacking an upper-bound does not obtain in larger type spaces such as the one in Chapter 4, whereas the noise-free model with functional pressure in that chapter robustly leads to this outcome. What is more, it is unclear what role noisy perception should play in the selection of underspecified meaning. These results should therefore be taken as suggestive but not indicative of a relationship between the two. Our aim here was mainly conceptual and technical in nature. In the context of this investigation, they serve to underscore the discussion in Chapter 6.

Bibliography

Christopher Ahern and Robin Clark. Diachronic processes in language as signaling under conflicting interests. In Erica A. Cartmill, Sean Roberts, Heidi Lyn, and Hannah Cornish, editors, *The Evolution of Language: Proceedings of the 10th International Conference*, pages 25–32. World Scientific Press, 2014. doi: 10.1142/9789814603638_0002.

Christopher Ahern and Robin Clark. Conflict, cheap talk, and Jespersen's cycle. *Semantics and Pragmatics*, 10, 2017. doi: 10.3765/sp.10.11.

Nicholas Allott. Game theory and communication. In Anton Benz, Gerhard Jäger, and Robert van Rooij, editors, *Game Theory and Pragmatics*, pages 123–152. Palgrave Macmillan, 2006. doi: 10.1057/9780230285897_4.

John Robert Anderson. *The Adaptive Character of Thought.* Psychology Press, 1990.

Kate Arnold and Klaus Zuberbühler. The alarm-calling system of adult male putty-nosed monkeys, cercopithecus nictitans martini. *Animal Behaviour*, 72 (3):643–653, 2006. doi: 10.1016/j.anbehav.2005.11.017.

Johan van der Auwera. On the diachrony of negation. In *The Expression of Negation*, pages 73–110. Walter de Gruyter GmbH, 2010. doi: 10.1515/9783110219302.73.

Ellen Gurman Bard, Anne H. Anderson, Catherine Sotillo, Matthew Aylett, Gwyneth Doherty-Sneddon, and Alison Newlands. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42(1):1–22, 2000. doi: 10.1006/jmla.1999.2667.

Andrea Baronchelli, Andrea Puglisi, and Vittorio Loreto. Cultural route to the emergence of linguistic categories. *PNAS*, 105(23):7936–7940, 2008. doi: 10.1073/pnas.0802485105.

Jeffrey Barrett and Kevin J.S. Zollman. The role of forgetting in the evolution and learning of language. *Journal of Experimental & Theoretical Artificial Intelligence*, 21(4):293–309, 2009. doi: 10.1080/09528130902823656.

John Batali. Computational simulations of the emergence of grammar. In James R. Hurford, Michael Studdert-Kennedy, and Chris Knight, editors, *Evolution of Language: Social and Cognitive Bases*. Cambridge University Press, Cambridge, UK, 1998.

A.W. Beggs. On the convergence of reinforcement learning. *Journal of Economic Theory*, 122(1):1–36, 2005. doi: 10.1016/j.jet.2004.03.008.

Anton Benz. Utility and relevance of answers. In Anton Benz, Gerhard Jäger, and Robert van Rooij, editors, *Game Theory and Pragmatics*, pages 195–219. Palgrave Macmillan, 2006. doi: 10.1057/9780230285897_7.

Anton Benz and Robert van Rooij. Optimal assertions, and what they implicate. A uniform game theoretic approach. *Topoi*, 26(1):63–78, 2007. doi: 10.1007/s11245-006-9007-3.

Anton Benz, Gerhard Jäger, and Robert van Rooij, editors. *Game Theory and Pragmatics*. Palgrave Macmillan, 2006a.

Anton Benz, Gerhard Jäger, and Robert van Rooij. An introduction to game theory for linguists. In Anton Benz, Gerhard Jäger, and Robert van Rooij, editors, *Game Theory and Pragmatics*, pages 1–82. Palgrave Macmillan, 2006b.

Leon Bergen and Noah D. Goodman. The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science*, 7(2):336–350, 2015. doi: 10.1111/tops.12144.

Leon Bergen, Noah D. Goodman, and Roger Levy. That's what she (could have) said: How alternative utterances affect language use. In *Proceedings of 34th Annual Meeting of the Cognitive Science Society*, 2012.

Leon Bergen, Roger Levy, and Noah D. Goodman. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9, 2016. doi: 10.3765/sp.9.20.

B. Douglas Bernheim. Rationalizable strategic behavior. *Econometrica*, 52(4):1007–1028, 1984. doi: 10.2307/1911196.

Ellen Bialystok, Raluca Barac, Agnes Blaye, and Diane Poulin-Dubois. Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognition and Development*, 11(4):485–508, 2010. doi: 10.1080/15248372.2010.516420.

Ricardo A.H. Bion, Arielle Borovsky, and Anne Fernald. Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30months. *Cognition*, 126(1):39–53, 2013. doi: 10.1016/j.cognition.2012.08.008.

Andreas Blume and Oliver Board. Intentional vagueness. *Erkenntnis*, 79(S4): 855–899, 2014. doi: 10.1007/s10670-013-9468-x.

Andreas Blume, Yong-Gwan Kim, and Joel Sobel. Evolutionary stability in games of communication. *Games and Economic Behavior*, 5(4):547–575, 1993. doi: 10.1006/game.1993.1031.

Johan J. Bolhuis, Ian Tattersall, Noam Chomsky, and Robert C. Berwick. How could language have evolved? *PLoS Biology*, 12(8):e1001934, 2014. doi: 10.1371/journal.pbio.1001934.

Dwight Bolinger. The deflation of several. *Journal of English Linguistics*, 15(1): 1–3, 1981.

Tilman Börgers and Rajiv Sarin. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, 77(1):1–14, 1997. doi: 10.1006/jeth.1997.2319.

Lewis Bott and Ira A. Noveck. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 51(3): 437–457, 2004. doi: 10.1016/j.jml.2004.05.006.

Melissa Bowerman. *Linguistic Typology and First Language Acquisition*. Oxford University Press, 2010. doi: 10.1093/oxfordhb/9780199281251.013.0028.

Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. Linguistic alignment between people and computers. *Journal of Pragmatics*, 42(9):2355–2368, 2010. doi: 10.1016/j.pragma.2009.12.012.

Richard Breheny, Napoleon Katsos, and John Williams. Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, 100(3):434–463, 2006. doi: 10.1016/j.cognition.2005.07.003.

Susan E. Brennan and Herbert H. Clark. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482–1493, 1996. doi: 10.1037/0278-7393.22.6.1482.

Susan E. Brennan and Joy E. Hanna. Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1(2):274–291, 2009. doi: 10.1111/j.1756-8765.2009.01019.x.

Henry Brighton. Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54, 2002. doi: 10.1162/106454602753694756.

Thomas Brochhagen. Minimal requirements for productive compositional signaling. In D.C. Noelle, R. Dale, A.S. Warlaumont, J. Yoshimi, T. Matlock, C.D. Jennings, and P.P. Maglio, editors, *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, pages 285–290. Cognitive Science Society, 2015a.

Thomas Brochhagen. Improving coordination on novel meaning through context and semantic structure. In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*, pages 74–82, 2015b.

Thomas Brochhagen. Signaling under uncertainty: Interpretative alignment without a common prior. *The British Journal for the Philosophy of Science*, 2017. doi: 10.1093/bjps/axx058.

Thomas Brochhagen and Michael Franke. Effects of transmission perturbation in the cultural evolution of language. In G. Gunzelmann, A. Howes, T. Tenbrink, and E.J. Davelaar, editors, *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, pages 1678–1683. Cognitive Science Society, 2017.

Thomas Brochhagen, Michael Franke, and Robert van Rooij. Learning biases may prevent lexicalization of pragmatic inferences: a case study combining iterated (Bayesian) learning and functional selection. In A. Papafragou, D. Grodner, D. Mirman, and J.C. Trueswell, editors, *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, pages 2081–2086. Cognitive Science Society, 2016.

Thomas Brochhagen, Michael Franke, and Robert van Rooij. Co-evolution of lexical meaning and pragmatic use, manuscript.

David Burkett and Thomas L. Griffiths. Iterated learning of multiple languages from multiple teachers. In A.D.M. Smith, M. Schouwstra, B. de Boer, and K. Smith, editors, *The Evolution of Language: Proceedings of the 8th international conference*, pages 58–65. World Scientific, 2010. doi: 10.1142/9789814295222_0008.

C.F. Camerer, T.-H. Ho, and J.-K. Chong. A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119(3):861–898, 2004. doi: 10.1162/0033553041502225.

Robyn Carston. Relevance theory and the saying/implicating distinction. In *The Handbook of Pragmatics*, pages 633–656. Blackwell Publishing Ltd, 2006. doi: 10.1002/9780470756959.ch28.

David Catteeuw and Bernard Manderick. The limits and robustness of reinforcement learning in Lewis signalling games. *Connection Science*, 26(2):161–177, 2014. doi: 10.1080/09540091.2014.885303.

Nick Chater and Paul Vitányi. Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences*, 7(1):19–22, 2003. doi: 10.1016/s1364-6613(02)00005-0.

Dorothy L. Cheney and Robert M. Seyfarth. *How monkeys see the world: Inside the mind of another species.* University of Chicago Press, 1990.

Gennaro Chierchia, Danny Fox, and Benjamin Spector. Scalar implicature as a grammatical phenomenon. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors, *Semantics. An International Handbook of Natural Language Meaning*, pages 2297–2332. de Gruyter, Berlin, 2012.

Noam Chomsky. *On Nature and Language.* Cambridge University Press, 2002.

Noam Chomsky. On phases. In *Foundational Issues in Linguistic Theory*, pages 132–166. MIT Press, 2008. doi: 10.7551/mitpress/9780262062787.003.0007.

Morten H. Christiansen and Nick Chater. Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(05), 2008. doi: 10.1017/s0140525x08004998.

Eve V. Clark. Lexical meaning. In Edith L. Bavin, editor, *The Cambridge Handbook of Child Language*, pages 283–300. Cambridge University Press, 2009.

Herbert H. Clark and Michael F. Schober. Asking questions and influencing answers. In *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, pages 14–48, 1992.

Herbert H. Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22(1):1–39, 1986. doi: 10.1016/0010-0277(86)90010-7.

Hannah Cornish. *Language adapts: Exploring the cultural dynamics of iterated learning.* PhD thesis, University of Edinburgh, 2011.

Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6):1431, 1982. doi: 10.2307/1913390.

Ross Cressman. *Evolutionary dynamics and extensive form games.* MIT Press, 2003.

Östen Dahl. Typology of sentence negation. *Linguistics*, 17(1-2), 1979.

Marcel Danesi. *Vico, Metaphor, and the Origin of Language.* Indiana University Press, 1993.

Isabelle Dautriche. *Weaving an Ambiguous Lexicon.* PhD thesis, École Normale Supérieure, 2015.

Isabelle Dautriche, Kyle Mahowald, Edward Gibson, Anne Christophe, and Steven T. Piantadosi. Words cluster phonetically beyond phonotactic regularities. *Cognition*, 163:128–145, 2017. doi: 10.1016/j.cognition.2017.02.001.

Kris De Jaegher and Robert van Rooij. Strategic vagueness, and appropriate contexts. In *Language, Games, and Evolution*, pages 40–59. Springer Science+Business Media, 2011. doi: 10.1007/978-3-642-18006-4_3.

Kris De Jaegher and Robert van Rooij. Game-theoretic pragmatics under conflicting and common interests. *Erkenntnis*, 2014. doi: 10.1007/s10670-013-9465-0.

Wim De Neys and Walter Schaeken. When people are more logical under cognitive load. *Experimental Psychology*, 54(2):128–133, 2007. doi: 10.1027/1618-3169.54.2.128.

Judith Degen and Michael Franke. Optimal reasoning about referential expressions. In Sarah Brown-Schmidt, Jonathan Ginzburg, and Staffan Larsoon, editors, *Proceedings of SemDial (SeineDial)*, pages 2–11, 2012.

Judith Degen and Michael K. Tanenhaus. Processing scalar implicatures: A constraint-based approach. *Cognitive Science*, 39:667–710, 2015. doi: 10.1111/cogs.12171.

Ashwini Deo. The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift. *Semantics & Pragmatics*, 8(1): 1–52, 2015. doi: 10.3765/sp.8.14.

Ido Erev and Alvin E. Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 88(4):848–881, 1998.

Joseph Farrell. Meaning and credibility in cheap-talk games. *Games and Economic Behavior*, 5(4):514–531, 1993. doi: 10.1006/game.1993.1029.

Nicolas Fay and T. Mark Ellison. The cultural evolution of human communication systems in different sized populations: Usability trumps learnability. *PLoS ONE*, 8(8), 2013. doi: 10.1371/journal.pone.0071781.

Jacob Feldman. Minimization of Boolean complexity in human concept learning. *Nature*, 407(6804):630–633, 2000.

Vanessa Ferdinand and Willem Zuidema. Thomas' theorem meets Bayes' rule: A model of the iterated learning of language. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, pages 2974–2979, 2009.

Victor S. Ferreira. Ambiguity, accessibility, and a division of labor for communicative success. In *Psychology of Learning and Motivation: Advances in Research and Theory*, pages 209–246. Elsevier BV, 2008. doi: 10.1016/s0079-7421(08)00006-6.

Victor S. Ferreira, L. Robert Slevc, and Erin S. Rogers. How do speakers avoid ambiguous linguistic expressions? *Cognition*, 96(3):263–284, 2005. doi: 10.1016/j.cognition.2004.09.002.

Francesca Foppolo and Marco Marelli. No delay for some inferences. *Journal of Semantics*, pages 1–23, 2017. doi: 10.1093/jos/ffx013.

Carol A. Fowler and Jonathan Housum. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26(5):489 – 504, 1987. doi: http://dx.doi.org/10.1016/0749-596X(87)90136-7.

Michael C. Frank and Noah D. Goodman. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998, 2012. doi: 10.1126/science.1218633.

Michael Franke. Interpretation of optimal signals. In Krzysztof R. Apt and Robert van Rooij, editors, *New Perspectives on Games and Interaction*, volume 4 of *Texts in Logic and Games*, pages 297–310. Amsterdam University Press, 2008.

Michael Franke. *Signal to Act: Game Theoretic Pragmatics*. PhD thesis, University of Amsterdam, 2009.

Michael Franke. Game theoretic pragmatics. *Philosophy Compass*, 8(3):269–284, 2013. doi: 10.1111/phc3.12015.

Michael Franke. The evolution of compositionality in signaling games. *Journal of Logic, Language and Information*, 25(3–4):355–377, 2016. doi: 10.1007/s10849-015-9232-5.

Michael Franke and José Pedro Correia. Vagueness and imprecise imitation in signalling games. *The British Journal for the Philosophy of Science*, 2017. doi: 10.1093/bjps/axx002.

Michael Franke and Judith Degen. Reasoning in reference games: Individual- vs. population-level probabilistic modeling. *PLoS ONE*, 11(5), 2016. doi: 10.1371/journal.pone.0154854.

Michael Franke and Gerhard Jäger. Bidirectional optimization from reasoning and learning in games. *Journal of Logic, Language and Information*, 21(1):117–139, 2011. doi: 10.1007/s10849-011-9151-z.

Michael Franke and Gerhard Jäger. Pragmatic back-and-forth reasoning. In *Pragmatics, Semantics and the Case of Scalar Implicatures*. Nature Publishing Group, 2014. doi: 10.1057/9781137333285.0011.

Michael Franke and Gerhard Jäger. Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, 35 (1), 2016a. doi: 10.1515/zfs-2016-0002.

Michael Franke and Gerhard Jäger. Reply to commentaries. *Zeitschrift für Sprachwissenschaft*, 35(1):117 – 132, 2016b. doi: 10.1515/zfs-2016-0009.

Michael Franke and Elliott O. Wagner. Game theory and the evolution of meaning. *Language and Linguistics Compass*, 8(9):359–372, 2014. doi: 10.1111/lnc3.12086.

Michael Franke, Gerhard Jäger, and Robert van Rooij. Vagueness, signaling and bounded rationality. In T. Onoda, D. Bekki, and E. McCready, editors, *New Frontiers in Artificial Intelligence*, pages 45–59. Springer Science+Business Media, 2011. doi: 10.1007/978-3-642-25655-4_5.

R. Fusaroli, B. Bahrami, K. Olsen, A. Roepstorff, G. Rees, C. Frith, and K. Tylen. Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, 23(8):931–939, 2012. doi: 10.1177/0956797612436816.

Simon Garrod and Gwyneth Doherty. Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3):181–215, 1994. doi: 10.1016/0010-0277(94)90048-5.

Gerald Gazdar. *Pragmatics, Implicature, Presuposition and Logical Form.* Academic Press, New York, 1979.

Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992. doi: 10.1214/ss/1177011136.

Bart Geurts. *Quantity Implicatures.* Cambridge University Press, Cambridge, UK, 2010.

Tao Gong. *Language Evolution from a Simulation Perspective: On the Coevolution of Compositionality and Regularity.* PhD thesis, Chinese University of Hong Kong, 2007.

Noah D. Goodman and Michael C. Frank. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11):818–829, 2016. doi: 10.1016/j.tics.2016.08.005.

Noah D. Goodman and Andreas Stuhlmüller. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, 5:173–184, 2013. doi: 10.1111/tops.12007.

Noah D. Goodman, Joshua Tenenbaum, Jacob Feldman, and Thomas Griffiths. A rational analysis of rule-based concept learning. *Cognitive Science: A Multidisciplinary Journal*, 32(1):108–154, 2008. doi: 10.1080/03640210701802071.

Stephen Jay Gould. *Ontogeny and Phylogeny.* Harvard University Press, 1977.

Alison Greenough, Graham Cole, Jonathan Lewis, Andrew Lockton, and John Blundell. Untangling the effects of hunger, anxiety, and nausea on energy intake during intravenous cholecystokinin octapeptide (CCK-8) infusion. *Physiology & Behavior*, 65(2):303–310, 1998. doi: 10.1016/s0031-9384(98)00169-3.

Paul Grice. Logic and conversation. *Syntax and Semantics*, 3:41–58, 1975.

Paul Grice. *Studies in the Way of Words.* Harvard University Press, 1989.

Thomas L. Griffiths and Michael L. Kalish. A Bayesian view of language evolution by iterated learning. In Bruno G. Bara, Lawrence Barsalou, and Monica Bucciarelli, editors, *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, pages 827–832. Cognitive Science Society, 2005.

Thomas L. Griffiths and Michael L. Kalish. Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, 31(3):441–480, 2007. doi: 10.1080/15326900701326576.

Thomas L. Griffiths, Nick Chater, Dennis Norris, and Alexandre Pouget. How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012). *Psychological Bulletin*, 138(3):415–422, 2012. doi: 10.1037/a0026884.

Daniel J. Grodner, Natalie M. Klein, Kathleen M. Carbary, and Michael K. Tanenhaus. "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, 166:42–55, 2010. doi: 10.1016/j.cognition.2010.03.014.

Ernst Haeckel. *Generelle Morphologie der Organismen: Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte Descendenz-Theorie*, volume 2. Georg Reimer, Berlin, 1866.

Justin Halberda. The development of a word-learning strategy. *Cognition*, 87(1): B23–B34, 2003. doi: 10.1016/s0010-0277(02)00186-5.

Justin Halberda. Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, 53(4):310–344, 2006. doi: 10.1016/j.cogpsych.2006.04.003.

Robert X.D. Hawkins, Michael C. Frank, and Noah D. Goodman. Convention-formation in iterated reference games. In G. Gunzelmann, A. Howes, T. Tenbrink, and E.J. Davelaar, editors, *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, pages 482–487. Cognitive Science Society, 2017.

Dirk Helbing. A stochastic behavioral model and a 'microscopic' foundation of evolutionary game theory. *Theory and Decision*, 40(2):149–179, 1996. doi: 10.1007/BF00133171.

Richard J. Herrnstein. On the law of effect. *Journal of the Experimental Analysis of Behavior*, 13(2):243–266, 1970.

Julia Hirschberg. *A Theory of Scalar Implicature.* PhD thesis, University of Pennsylvania, 1985.

Charles F. Hockett. The origin of speech. *Scientific American*, (203):89–97, 1960.

Josef Hofbauer. The selection mutation equation. *Journal of Mathematical Biology*, 23:41–53, 1985. doi: 10.1007/bf00276557.

Josef Hofbauer and Karl Sigmund. Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, 40(04):479–520, 2003. doi: 10.1090/s0273-0979-03-00988-1.

Laurence R. Horn. *On the Semantic Properties of Logical Operators in English.* Indiana University Linguistics Club, Bloomington, IN, 1972.

Laurence R. Horn. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin, editor, *Meaning, Form and Use in Context*, pages 11 – 42. Georgetown University Press, 1984.

Carmel Houston-Price, Zoe Caloghiris, and Eleonora Raviglione. Language experience shapes the development of the mutual exclusivity bias. *Infancy*, 15(2): 125–150, 2010. doi: 10.1111/j.1532-7078.2009.00009.x.

Yi Ting Huang and Jesse Snedeker. Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, 58(3): 376–415, 2009.

Carla L. Hudson Kam and Elissa Newport. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2):151–195, 2005.

James R. Hurford. Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, 77(2):187–222, 1989.

Simon M. Huttegger. Evolution and the explanation of meaning. *Philosophy of Science*, 74(1):1–27, 2007. doi: 10.1086/519477.

Simon M. Huttegger and Kevin J.S. Zollman. Methodology in biological game theory. *The British Journal for the Philosophy of Science*, 64(3):637–658, 2013. doi: 10.1093/bjps/axs035.

Simon M. Huttegger, Brian Skyrms, and Kevin J.S. Zollman. Probe and adjust in information transfer games. *Erkenntnis*, 79(S4):835–853, 2013. doi: 10.1007/s10670-013-9467-y.

Nori Jacoby and Josh H. McDermott. Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*, 27(3):359–370, 2017. doi: 10.1016/j.cub.2016.12.031.

Gerhard Jäger. Evolutionary game theory and typology: A case study. *Language*, 83(1):74–109, 2007a. doi: 10.2307/4490338.

Gerhard Jäger. Game dynamics connects semantics and pragmatics. In Ahti-Veikko Pietarinen, editor, *Game Theory and Linguistic Meaning*, pages 89–102. Elsevier, 2007b.

Gerhard Jäger and Robert van Rooij. Language structure: psychological and social constraints. *Synthese*, 159(1):99–130, 2007. doi: 10.1007/s11229-006-9073-5.

Brendan Juba, Adam Tauman Kalai, Sanjeev Khanna, and Madhu Sudan. Compression without a common prior: An information-theoretic justification for ambiguity in language. In *Proceedings of the 2nd Symposium on innovations in computer science*, 2011.

Michael L. Kalish, Thomas L. Griffiths, and Stephan Lewandowsky. Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, 14(2):288–294, 2007. doi: 10.3758/bf03194066.

Jasmeen Kanwal, Kenny Smith, Jennifer Culbertson, and Simon Kirby. Zipf's law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, 165:45–52, 2017. doi: 10.1016/j.cognition.2017.05.001.

Yarden Katz, Noah D. Goodman, Kristian Kersting, Charles Kemp, and Joshua B. Tenenbaum. Modeling semantic cognition as logical dimensionality reduction. In *Proceedings of 30th Annual Meeting of the Cognitive Science Society*, 2008.

C. Kemp and T. Regier. Kinship categories across languages reflect general communicative principles. *Science*, 336(6084):1049–1054, 2012. doi: 10.1126/science.1218811.

Midam Kim, William S. Horton, and Ann R. Bradlow. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, 2(1), 2011. doi: 10.1515/labphon.2011.004.

Motoo Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, 1983. doi: 10.1017/cbo9780511623486.

Simon Kirby. Spontaneous evolution of linguistic structure - an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2):102–110, 2001. doi: 10.1109/4235.918430.

Simon Kirby. Learning, bottlenecks and the evolution of recursive syntax. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition*, pages 173–204. Cambridge University Press, 2002. doi: 10.1017/cbo9780511486524.006.

Simon Kirby and James R. Hurford. The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi and D. Parisi, editors, *Simulating the Evolution of Language*, pages 121–148. Springer, 2002.

Simon Kirby, M. Dowman, and Thomas L. Griffiths. Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(12):5241–5245, 2007. doi: 10.1073/pnas.0608222104.

Simon Kirby, H. Cornish, and K. Smith. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31):10681–10686, 2008. doi: 10.1073/pnas.0707835105.

Simon Kirby, Tom Griffiths, and Kenny Smith. Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28:108–114, 2014. doi: 10.1016/j.conb.2014.07.014.

Simon Kirby, Monica Tamariz, Hannah Cornish, and Kenny Smith. Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102, 2015. doi: 10.1016/j.cognition.2015.03.016.

Robert M. Krauss and Sidney Weinheimer. Changes in reference phrases as a function of frequency of usage in social interaction: a preliminary study. *Psychonomic Science*, 1(1-12):113–114, 1964. doi: 10.3758/bf03342817.

Daniel Lassiter and Noah D. Goodman. Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, 2015. doi: 10.1007/s11229-015-0786-1.

Michael D. Lee and Eric-Jan Wagenmakers. *Bayesian cognitive modeling: A practical course.* Cambridge university press, 2014.

Tom Lenaerts, Bart Jansen, Karl Tuyls, and Bart De Vylder. The evolutionary language game: An orthogonal approach. *Journal of Theoretical Biology*, 235: 566–582, 2005. doi: 10.1016/j.jtbi.2005.02.009.

Stephen C. Levinson. *Pragmatics.* Cambridge University Press, Cambridge, UK, 1983.

Stephen C. Levinson. *Presumptive Meanings: The Theory of Generalized Conversational Implicature.* MIT Press, 2000.

David Lewis. *Convention: A Philosophical Study.* Harvard University Press, Cambridge, 1969.

Barton L. Lipman. Why is language vague? Manuscript, Boston University, 2009.

Duncan R. Luce. *Individual choice behavior: a theoretical analysis.* Wiley, 1959.

Kyle Mahowald, Evelina Fedorenko, Steven T. Piantadosi, and Edward Gibson. Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2):313–318, 2013. doi: 10.1016/j.cognition.2012.09.010.

Ellen M. Markman and Gwyn F. Wachtel. Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, 20(2):121–157, 1988.

David Marr. *Vision.* MIT Press, 1982.

André Martinet. *A Functional View of Language.* Clarendon Press, Oxford, 1962.

Steven A. Matthews, Masahiro Okuno-Fujiwara, and Andrew Postlewaite. Refining cheap-talk equilibria. *Journal of Economic Theory*, 55(2):247–273, 1991. doi: 10.1016/0022-0531(91)90040-b.

J. Maynard Smith and G. R. Price. The logic of animal conflict. *Nature*, 246 (5427):15–18, 1973. doi: 10.1038/246015a0.

Judith Mehta, Chris Starmer, and Robert Sugden. Focal points in pure coordination games: An experimental investigation. *Theory and Decision*, 36(2): 163–185, 1994. doi: 10.1007/bf01079211.

William E. Merriman, Laura L. Bowman, and Brian MacWhinney. The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, 54(3), 1989. doi: 10.2307/1166130.

Charles Metzing and Susan E. Brennan. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49(2):201–213, 2003. doi: 10.1016/s0749-596x(03)00028-7.

George A. Miller. *Language and Communication.* McGraw-Hill, 1951.

Yasamin Motamedi, Marieke Schouwstra, Kenny Smith, and Simon Kirby. Linguistic structure emerges in the cultural evolution of artificial sign languages. In *The Evolution of Language: Proceedings of the 11th International Conference*, 2016.

John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, 1944.

Mante S. Nieuwland, Tali Ditman, and Gina R. Kuperberg. On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, 63(3):324–346, 2010. doi: 10.1016/j.jml.2010.06.005.

Martin A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life.* Harvard University Press, 2006.

Martin A. Nowak and D.C. Krakauer. The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14):8028–8033, 1999. doi: 10.1073/pnas.96.14.8028.

Martin A. Nowak, Joshua B. Plotkin, and Vincent A.A. Jansen. The evolution of syntactic communication. *Nature*, 404(6777):495–498, 2000. doi: 10.1038/35006635.

Martin A. Nowak, Natalia L. Komarova, and Partha Niyogi. Evolution of universal grammar. *Science*, 291(5501):114–118, 2001. doi: 10.1126/science.291.5501.114.

Martin A. Nowak, Natalia L. Komarova, and Partha Niyogi. Computational and evolutionary aspects of language. *Nature*, 417(6889):611–617, 2002. doi: 10.1038/nature00771.

Cailin O'Connor. The evolution of vagueness. *Erkenntnis*, 79(S4):707–727, 2014. doi: 10.1007/s10670-013-9463-2.

Cailin O'Connor. Ambiguity is kinda good sometimes. *Philosophy of Science*, 82 (1):pp. 110–121, 2015. doi: 10.1086/679180.

Martin J. Osborne. *An Introduction to Game Theory*. Oxford University Press, 2004.

Mark Pagel. Human language as a culturally transmitted replicator. *Nature Reviews Genetics*, 10:405–415, 2009. doi: 10.1038/nrg2560.

Prashant Parikh. Communication and strategic inference. *Linguistics and Philosophy*, 14(5):473–514, 1991. doi: 10.1007/bf00632595.

Prashant Parikh. A game-theoretic account of implicature. In Yoram Moses, editor, *Proceedings of the 4th conference on Theoretical aspects of reasoning about knowledge*, pages 85–94. Morgan Kaufmann Publishers Inc., 1992.

Prashant Parikh. Communication, meaning, and interpretation. *Linguistics and Philosophy*, 23(2):185–212, 2000.

Rohit Parikh. Vagueness and utility: The semantics of common nouns. *Linguistics and Philosophy*, 17(6):521–535, 1994. doi: 10.1007/bf00985317.

John K. Pate and Sharon Goldwater. Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language*, 78:1–17, 2015. doi: 10.1016/j.jml.2014.10.003.

Cédric Paternotte and Jonathan Grose. Robustness in evolutionary explanations: a positive account. *Biology & Philosophy*, 32(1):73–96, 2017. doi: 10.1007/s10539-016-9539-x.

Christina Pawlowitsch. Why evolution does not always lead to an optimal signaling system. *Games and Economic Behavior*, 63(1):203–226, 2008. doi: 10.1016/j.geb.2007.08.009.

David G. Pearce. Rationalizable strategic behavior and the problem of perfection. *Econometrica*, 52(4):1029, 1984. doi: 10.2307/1911197.

Amy Perfors and Daniel J. Navarro. Language evolution can be shaped by the structure of the world. *Cognitive Science*, 38(4):775–793, 2014. doi: 10.1111/cogs.12102.

Steven T. Piantadosi. Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review*, 21(5):1112–1130, 2014. doi: 10.3758/s13423-014-0585-6.

Steven T. Piantadosi and Robert A. Jacobs. Four problems solved by the probabilistic language of thought. *Current Directions in Psychological Science*, 25 (1):54–59, 2016. doi: 10.1177/0963721415609581.

Steven T. Piantadosi, H. Tily, and E. Gibson. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108 (9):3526–3529, 2011. doi: 10.1073/pnas.1012551108.

Steven T. Piantadosi, Joshua B. Tenenbaum, and Noah D. Goodman. Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*, 123(2):199–217, 2012a. doi: 10.1016/j.cognition.2011.11.005.

Steven T. Piantadosi, Harry Tily, and Edward Gibson. The communicative function of ambiguity in language. *Cognition*, 122(3):280–291, 2012b. doi: 10.1016/j.cognition.2011.10.004.

Steven T. Piantadosi, Joshua B. Tenenbaum, and Noah D. Goodman. Modeling the acquisition of quantifier semantics: a case study in function word learnability, under review.

Martin J. Pickering and Victor S. Ferreira. Structural priming: A critical review. *Psychological Bulletin*, 134(3):427–459, 2008. doi: 10.1037/0033-2909.134.3.427.

Martin J. Pickering and Simon Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02), 2004. doi: 10.1017/s0140525x04000056.

Tabitha Price, Philip Wadewitz, Dorothy Cheney, Robert Seyfarth, Kurt Hammerschmidt, and Julia Fischer. Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific Reports*, 5(1), 2015. doi: 10.1038/srep13220.

Ciyang Qing and Michael Franke. Variations on a Bayesian theme: Comparing Bayesian models of referential reasoning. In *Bayesian Natural Language Semantics and Pragmatics*, pages 201–220. Springer International Publishing, 2015. doi: 10.1007/978-3-319-17064-0_9.

Matthew Rabin. Communication between rational agents. *Journal of Economic Theory*, 51(1):144–170, 1990. doi: 10.1016/0022-0531(90)90055-o.

Florencia Reali and Thomas L. Griffiths. Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1680):429–436, 2009. doi: 10.1098/rspb.2009.1513.

David Reitter and Johanna D. Moore. Alignment and task success in spoken dialogue. *Journal of Memory and Language*, 76:29–46, 2014. doi: 10.1016/j.jml.2014.05.008.

Mark Ridley. Adaptation. In Mark Pagel, editor, *Encyclopedia of evolution*, volume 1, pages 10–15. Oxford University Press, 2002. doi: 0.1093/acref/9780195122008.001.0001.

Robert van Rooij and Katrin Schulz. Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information*, 13(4):491–519, 2004. doi: 10.1007/s10849-004-2118-6.

Robert van Rooij and Merljin Sevenster. Different faces of risky speech. In Anton Benz, Gerhard Jäger, and Robert van Rooij, editors, *Game Theory and Pragmatics*. Palgrave Macmillan, 2006.

Robert van Rooy. Signalling Games Select Horn Strategies. *Linguistics and Philosophy*, 27(4):493–527, 2004a. doi: 10.1023/b:ling.0000024403.88733.3f.

Robert van Rooy. Utility, informativity and protocols. *Journal of Philosophical Logic*, 33(4):389–419, 2004b. doi: 10.1023/b:logi.0000036830.62877.ee.

Robert van Rooij and Tikitu de Jager. Explaining quantity implicatures. *Journal of Logic, Language and Information*, 21(4):461–477, 2012. doi: 10.1007/s10849-012-9163-3.

Alvin E. Roth and Ido Erev. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8(1):164–212, 1995. doi: 10.1016/s0899-8256(05)80020-x.

John Salvatier, Thomas V. Wiecki, and Christopher Fonnesbeck. Probabilistic programming in python using PyMC3. *PeerJ Computer Science*, 2:e55, 2016. doi: 10.7717/peerj-cs.55.

William H. Sandholm. *Population Games and Evolutionary Dynamics.* MIT Press, Cambridge, MA, 2010.

Carlos Santana. Ambiguity in cooperative signaling. *Philosophy of Science*, 81 (3):398–422, 2014. doi: 10.1086/676652.

Uli Sauerland. Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27:367–391, 2004. doi: 10.1023/B:LING.0000023378.71748.db.

Gerhard Schaden. Modelling the “aoristic drift of the present perfect” as inflation: An essay in historical pragmatics. *International Review of Pragmatics*, 4:261–292, 2012.

Thomas C. Schelling. *The Strategy of Conflict*. Harvard University Press, 1960.

Karl H. Schlag. Why imitate, and if so, how? *Journal of Economic Theory*, 78 (1):130–156, 1998. doi: doi:10.1006/jeth.1997.2347.

William A. Searcy and Stephen Nowicki. *The evolution of animal communication: reliability and deception in signaling systems*. Princeton University Press, 2005.

John R. Searle. What is a speech act? In John R. Searle, editor, *The Philosophy of Language*. Oxford University Press, 1971.

Robert M. Seyfarth, Dorothy L. Cheney, and Peter Marler. Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour*, 28(4):1070–1094, 1980. doi: 10.1016/s0003-3472(80)80097-2.

Catriona Silvey, Simon Kirby, and Kenny Smith. Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive Science*, 39(1): 212–226, 2014. doi: 10.1111/cogs.12150.

Herbert A. Simon. Invariants of human behavior. *Annual Review of Psychology*, 41(1):1–19, 1990. doi: 10.1146/annurev.psych.41.1.1.

Brian Skyrms. *Signals: Evolution, learning, and information*. Oxford University Press, 2010.

Kenny Smith. Learning bias, cultural evolution of language, and the biological evolution of the language faculty. *Human Biology*, 83(2):261–278, 2011.

Kenny Smith, Simon Kirby, and Henry Brighton. Iterated learning: A framework for the emergence of language. *Artificial Life*, 9:371–386, 2003.

Dan Sperber and Deirdre Wilson. *Relevance: Communication and Cognition*. Wiley-Blackwell, 1986.

Matthew Spike, Kevin Stadler, Simon Kirby, and Kenny Smith. Minimal requirements for the emergence of learned signaling. *Cognitive Science*, pages 1–36, 2016. doi: 10.1111/cogs.12351.

Robert Stalnaker. Saying and meaning, cheap talk and credibility. In Anton Benz, Gerhard Jäger, and Robert van Rooij, editors, *Game Theory and Pragmatics*, pages 83–100. Palgrave Macmillan, 2006. doi: 10.1057/9780230285897_2.

Luc Steels. A self-organizing spatial vocabulary. *Artificial Life*, 2(3):319–332, 1995. doi: 10.1162/artl.1995.2.3.319.

Luc Steels. The origins of ontologies and communication conventions in multi-agent systems. *Agents and Multi-Agent Systems*, 1(2):169–194, 1998.

Luc Steels. Modeling the cultural evolution of language. *Physics of Life Reviews*, 8(4):339–356, 2011.

Luc Steels and Tony Belpaeme. Coordinating perceptually grounded categories through language: A case study for color. *Behavioral and Brain Sciences*, 28 (4):469–529, 2005. doi: 10.1017/S0140525X05000087.

Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning.* MIT Press, Cambridge, MA, USA, 1998.

Monica Tamariz and Simon Kirby. The cultural evolution of language. *Current Opinion in Psychology*, 8:37–43, 2016.

Peter D. Taylor and Leo B. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Bioscience*, 40(1–2):145–156, 1978.

Michael Henry Tessler and Noah D. Goodman. A pragmatic theory of generic language. Manuscript, Stanford University, arXiv:1608.02926, 2016.

Bill Thompson, Simon Kirby, and Kenny Smith. Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences of the United States of America*, 113(16):4530–4535, 2016. doi: 10.1073/pnas.1523631113.

Edward L. Thorndike. Animal intelligence: An experimental study of the associative processes in animals. *Psychological Monographs: General and Applied*, 2(4):i–109, 1898.

Bob van Tiel, Emiel van Miltenburg, Natalia Zevakhina, and Bart Geurts. Scalar diversity. *Journal of Semantics*, 33(1):137–175, 2014. doi: 10.1093/jos/ffu017.

John M. Tomlinson Jr., Todd M. Bailey, and Lewis Bott. Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, 69(1):18–35, 2013. doi: 10.1016/j.jml.2013.02.003.

Elizabeth Closs Traugott. Historical pragmatics. In Laurence R. Horn and Gregory Wand, editors, *The Handbook of Pragmatics*, pages 538–561. Blackwell Publishing, 2004.

Kees van Deemter. Utility and language generation: The case of vagueness. *Journal of Philosophical Logic*, 38(6):607–632, 2009.

Tessa Verhoef, Simon Kirby, and Bart de Boer. Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43:57–68, 2014. doi: 10.1016/j.wocn.2014.02.005.

Karl Wärneryd. Cheap talk, coordination, and evolutionary stability. *Games and Economic Behavior*, 5(4):532–546, 1993. doi: 10.1006/game.1993.1030.

Thomas Wasow. Ambiguity avoidance is overrated. In *Ambiguity*. Walter de Gruyter GmbH, 2015. doi: 10.1515/9783110403589-003.

Matthijs Westera. *Exhaustivity and Intonation: A Unified Theory*. PhD thesis, University of Amsterdam, 2017.

Timothy Williamson. *Vagueness*. Routledge, London and New York, 1994.

Deirdre Wilson and Dan Sperber. Relevance theory. In *The Handbook of Pragmatics*, pages 606–632. Wiley-Blackwell, 2006. doi: 10.1002/9780470756959.ch27.

Marieke Woensdregt and Kenny Smith. Pragmatics and language evolution. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, 2017. doi: 10.1093/acrefore/9780199384655.013.321.

James Woodward. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, 2004. doi: 10.1093/0195155270.001.0001.

Petri Ylikoski and N. Emrah Aydinonat. Understanding with theoretical models. *Journal of Economic Methodology*, 21(1):19–36, 2014. doi: 10.1080/1350178x.2014.886470.

E.J. Yoon, M.H. Tessler, N.D. Goodman, and M.C. Frank. Talking with tact: Polite language as a balance between kindness and informativity. In A. Papafragou, D. Grodner, D. Mirman, and J.C. Trueswell, editors, *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, pages 2771–2776. Cognitive Science Society, 2016.

George Zipf. *Human behavior and the principle of least effort*. Addison-Wesley Press, 1949.

Klaus Zuberbühler. Survivor signals: The biology and psychology of animal alarm calling. In *Advances in the Study of Behavior*, pages 277–322. Elsevier, 2009. doi: 10.1016/s0065-3454(09)40008-1.

Samenvatting

Wat wordt overgebracht gaat vaak verder dan wat er wordt gezegd. In plaats van het te vermijden, floreert alledaagse communicatie in het impliciete; in het onuitgesprokene; in het contextueel bepaalde. Dit onderzoek concentreert zich op deze kwestie door te vragen waarom, en onder welke voorwaarden, taal(gebruik) het onuitgesprokene ter hand neemt terwijl de zaken ook explicieter overgebracht hadden kunnen worden. Preciezer gezegd proberen we op een fundamenteel niveau te begrijpen waarom het werk verdeeld wordt tussen semantiek en pragmatiek. Dit doen we door te kijken naar de voorwaarden waaronder eigenschappen die van deze werkverdeling gebruik maken voortkomen. We analyseren deze voorwaarden door speltheoretische modellen van rationeel taalgebruik, *reinforcement learning*, (geïtereerd) Bayesiaans leren, en populatiedynamiek zoals de repliceerder-muteerderdynamiek op nieuwe manieren te combineren.

Onze analyse traceert taalverandering zowel op het niveau van geïtereerde interacties als op het populatieniveau. Beide niveaus brengen hun eigen perspectief mee en schijnen daarmee hun eigen licht op een gegeven eigenschap van taal. Dit maakt het mogelijk om verschillende, doch verweven, antwoorden op vragen zoals waarom alledaagse communicatie doorspekt is met semantische ambiguïteit; onder welke voorwaarden pragmatische inferenties mogelijk (niet) lexicalizeren; en, meer algemeen, wat voor werkverdelingen tussen semantiek en pragmatiek we kunnen verwachten voort te zien komen uit het krachtenveld en omgevingsfactoren die taal vormgeven te verkennen.

Op het niveau van geïtereerde interacties, analyseren we het opzettelijk gebruik van ambigue uitdrukkingen in dialoog. Aan de hand van eerdere verklaringen van ambiguïteit beargumenteren we dat context een belangrijke rol speelt in het mogelijk maken van de risicoloze uitbuiting van ambiguïteit. We slaan echter wat gaten in deze uitleg door de aanname dat dialoogpartners toegang hebben tot dezelfde contextuele informatie te betwijfelen en uiteindelijk op te geven. Deze kwestie ontvouwt zich in een grotere, waar het samenspel tussen context, de gesprekspartners' subjectieve contextuele verwachtingen, en hun opvattingen over elkaars verwachtingen een belangrijke rol spelen. We beredeneren dat de gezamenlijke uitkomst van deze factoren de voorwaarden bepaalt waaronder een functioneel voordeel voor ambiguïteit kristaliseert. Om deze ideeën tastbaar te maken stellen we een model van rationeel taalgebruik voor en koppelen het model aan simpele adaptieve dynamieken. We laten zien dat het model empirisch onderbouwde patronen van ambigu taalgebruik met succes voorspelt.

Op het populatieniveau staan eigenschappen emergent aan interacties aan de semantiek-pragmatiek interface voor uitdagingen die niet alleen taalgebruik, maar ook hun getrouwe overdracht over generaties omvatten. Nog semantiek nog pragmatiek zijn direct observeerbaar. In plaats daarvan zien leerlingen alleen het gedrag waarin de combinatie van semantiek en pragmatiek resulteert. Dit levert een probleem op omdat verschillende combinaties kunnen resulteren in (bijna) niet te onderscheiden zichtbare gedrag. In preciezere zin vragen we wanneer en waarom reguliere pragmatische inferenties wel (of niet) lexicalizeren, en wanneer semantische onderspecificatie ofwel wordt behouden ofwel plaatsmaakt voor preciezere uitdrukkingen. Om deze vragen aan te pakken formuleren we een model van de (co)evolutie van semantiek en pragmatiek. Het model houdt de effecten bij van functionele druk richting efficiënte informatieoverdracht en de effecten van druk richting leerbaarheid op zowel aparte als gecombineerde evolutionaire trajecten. We combineren dit model met modellen van pragmatisch taalgebruik op het niveau van het individu en dompelen het model onder in verschillende taalgebruik- en leeromgevingen.

Abstract

What is conveyed often goes beyond what is said. Rather than avoiding it, natural communication seems to thrive in the implicit; in the unsaid; in the contextually determined. This investigation centers around this issue by asking why and under which conditions language (use) may come to leverage or accommodate the unsaid when matters could be conveyed more explicitly. More precisely, at a fundamental level, we seek to better understand why there is a division of labor between semantics and pragmatics. We do so by looking at the conditions under which properties that draw from this division arise, which we analyze by combining, in novel ways, game-theoretic models of rational language use, reinforcement learning, (iterated) Bayesian learning, and population dynamics such as the replicator-mutator dynamic.

Our analysis traces linguistic change at the level of iterated interactions as well as at that of populations. Both levels come with their own perspective and thereby shed their own light on a given linguistic property. This allows us to explore different yet connected answers to questions such as why natural communication is rife with semantic ambiguity; under which conditions systematic pragmatic inferences may (fail to) lexicalize; and, more generally, what kinds of divisions of labor between semantics and pragmatics we can expect to arise from pressures and environmental factors that shape language.

At the level of iterated interactions, we analyze the deliberate use of ambiguous expressions in dialog. With previous explanations of ambiguity, we argue that context plays an important role in allowing for the safe exploitation of ambiguity. However, we inject some wrinkles into this explanation by calling into question and giving up the assumption that interlocutors have access to the same contextual information. This issue unravels into a larger one, where the interplay between context, interlocutors' subjective contextual expectations, and their beliefs about each other's expectations play an important role. We argue that the joint outcome of these factors determines the conditions under which a functional advantage for ambiguity crystallizes. We propose a model of rational language use and couple it with simple adaptive dynamics to capture these ideas, and show that it succeeds in predicting empirically attested patterns of ambiguous language use.

At the population level, properties that draw from interactions at the semantics-pragmatics interface face challenges not only in language use, but also in their faithful transmission across generations. Neither semantics nor pragmatics are directly observable. Learners instead only witness the behavior in which their combination results. This raises an issue because different divisions could result in (almost) indistinguishable overt linguistic behavior. More precisely, we ask why and when regular pragmatic inferences do (not) lexicalize, and when semantic underspecification is either maintained or gives way to more precise expressions. To address these questions, we formulate a model of the (co-)evolution of semantics and pragmatics. This model tracks the effects of functional pressure toward efficient information transfer and the effects of pressure for learnability on separate as well as on combined evolutionary trajectories. We combine this model with individual-level models of pragmatic language use, and couch it in different environments of language use and learning.

Titles in the ILLC Dissertation Series:

ILLC DS-2009-01: **Jakub Szymanik**
Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language

ILLC DS-2009-02: **Hartmut Fitz**
Neural Syntax

ILLC DS-2009-03: **Brian Thomas Semmes**
A Game for the Borel Functions

ILLC DS-2009-04: **Sara L. Uckelman**
Modalities in Medieval Logic

ILLC DS-2009-05: **Andreas Witzel**
Knowledge and Games: Theory and Implementation

ILLC DS-2009-06: **Chantal Bax**
Subjectivity after Wittgenstein. Wittgenstein's embodied and embedded subject and the debate about the death of man.

ILLC DS-2009-07: **Kata Balogh**
Theme with Variations. A Context-based Analysis of Focus

ILLC DS-2009-08: **Tomohiro Hoshi**
Epistemic Dynamics and Protocol Information

ILLC DS-2009-09: **Olivia Ladinig**
Temporal expectations and their violations

ILLC DS-2009-10: **Tikitu de Jager**
"Now that you mention it, I wonder…": Awareness, Attention, Assumption

ILLC DS-2009-11: **Michael Franke**
Signal to Act: Game Theory in Pragmatics

ILLC DS-2009-12: **Joel Uckelman**
More Than the Sum of Its Parts: Compact Preference Representation Over Combinatorial Domains

ILLC DS-2009-13: **Stefan Bold**
Cardinals as Ultrapowers. A Canonical Measure Analysis under the Axiom of Determinacy.

ILLC DS-2010-01: **Reut Tsarfaty**
Relational-Realizational Parsing

ILLC DS-2010-02: **Jonathan Zvesper**
Playing with Information

ILLC DS-2010-03: **Cédric Dégremont**
The Temporal Mind. Observations on the logic of belief change in interactive systems

ILLC DS-2010-04: **Daisuke Ikegami**
Games in Set Theory and Logic

ILLC DS-2010-05: **Jarmo Kontinen**
Coherence and Complexity in Fragments of Dependence Logic

ILLC DS-2010-06: **Yanjing Wang**
Epistemic Modelling and Protocol Dynamics

ILLC DS-2010-07: **Marc Staudacher**
Use theories of meaning between conventions and social norms

ILLC DS-2010-08: **Amélie Gheerbrant**
Fixed-Point Logics on Trees

ILLC DS-2010-09: **Gaëlle Fontaine**
Modal Fixpoint Logic: Some Model Theoretic Questions

ILLC DS-2010-10: **Jacob Vosmaer**
Logic, Algebra and Topology. Investigations into canonical extensions, duality theory and point-free topology.

ILLC DS-2010-11: **Nina Gierasimczuk**
Knowing One's Limits. Logical Analysis of Inductive Inference

ILLC DS-2010-12: **Martin Mose Bentzen**
Stit, Iit, and Deontic Logic for Action Types

ILLC DS-2011-01: **Wouter M. Koolen**
Combining Strategies Efficiently: High-Quality Decisions from Conflicting Advice

ILLC DS-2011-02: **Fernando Raymundo Velazquez-Quesada**
Small steps in dynamics of information

ILLC DS-2011-03: **Marijn Koolen**
The Meaning of Structure: the Value of Link Evidence for Information Retrieval

ILLC DS-2011-04: **Junte Zhang**
System Evaluation of Archival Description and Access

ILLC DS-2011-05: **Lauri Keskinen**
Characterizing All Models in Infinite Cardinalities

ILLC DS-2011-06: **Rianne Kaptein**
Effective Focused Retrieval by Exploiting Query Context and Document Structure

ILLC DS-2011-07: **Jop Briët**
Grothendieck Inequalities, Nonlocal Games and Optimization

ILLC DS-2011-08: **Stefan Minica**
Dynamic Logic of Questions

ILLC DS-2011-09: **Raul Andres Leal**
Modalities Through the Looking Glass: A study on coalgebraic modal logic and their applications

ILLC DS-2011-10: **Lena Kurzen**
Complexity in Interaction

ILLC DS-2011-11: **Gideon Borensztajn**
The neural basis of structure in language

ILLC DS-2012-01: **Federico Sangati**
Decomposing and Regenerating Syntactic Trees

ILLC DS-2012-02: **Markos Mylonakis**
Learning the Latent Structure of Translation

ILLC DS-2012-03: **Edgar José Andrade Lotero**
Models of Language: Towards a practice-based account of information in natural language

ILLC DS-2012-04: **Yurii Khomskii**
Regularity Properties and Definability in the Real Number Continuum: idealized forcing, polarized partitions, Hausdorff gaps and mad families in the projective hierarchy.

ILLC DS-2012-05: **David García Soriano**
Query-Efficient Computation in Property Testing and Learning Theory

ILLC DS-2012-06: **Dimitris Gakis**
Contextual Metaphilosophy - The Case of Wittgenstein

ILLC DS-2012-07: **Pietro Galliani**
The Dynamics of Imperfect Information

ILLC DS-2012-08: **Umberto Grandi**
Binary Aggregation with Integrity Constraints

ILLC DS-2012-09: **Wesley Halcrow Holliday**
Knowing What Follows: Epistemic Closure and Epistemic Logic

ILLC DS-2012-10: **Jeremy Meyers**
Locations, Bodies, and Sets: A model theoretic investigation into nominalistic mereologies

ILLC DS-2012-11: **Floor Sietsma**
Logics of Communication and Knowledge

ILLC DS-2012-12: **Joris Dormans**
Engineering emergence: applied theory for game design

ILLC DS-2013-01: **Simon Pauw**
Size Matters: Grounding Quantifiers in Spatial Perception

ILLC DS-2013-02: **Virginie Fiutek**
Playing with Knowledge and Belief

ILLC DS-2013-03: **Giannicola Scarpa**
Quantum entanglement in non-local games, graph parameters and zero-error information theory

ILLC DS-2014-01: **Machiel Keestra**
Sculpting the Space of Actions. Explaining Human Action by Integrating Intentions and Mechanisms

ILLC DS-2014-02: **Thomas Icard**
The Algorithmic Mind: A Study of Inference in Action

ILLC DS-2014-03: **Harald A. Bastiaanse**
Very, Many, Small, Penguins

ILLC DS-2014-04: **Ben Rodenhäuser**
A Matter of Trust: Dynamic Attitudes in Epistemic Logic

ILLC DS-2015-01: **María Inés Crespo**
Affecting Meaning. Subjectivity and evaluativity in gradable adjectives.

ILLC DS-2015-02: **Mathias Winther Madsen**
The Kid, the Clerk, and the Gambler - Critical Studies in Statistics and Cognitive Science

ILLC DS-2015-03: **Shengyang Zhong**
Orthogonality and Quantum Geometry: Towards a Relational Reconstruction of Quantum Theory

ILLC DS-2015-04: **Sumit Sourabh**
Correspondence and Canonicity in Non-Classical Logic

ILLC DS-2015-05: **Facundo Carreiro**
Fragments of Fixpoint Logics: Automata and Expressiveness

ILLC DS-2016-01: **Ivano A. Ciardelli**
Questions in Logic

ILLC DS-2016-02: **Zoé Christoff**
Dynamic Logics of Networks: Information Flow and the Spread of Opinion

ILLC DS-2016-03: **Fleur Leonie Bouwer**
What do we need to hear a beat? The influence of attention, musical abilities, and accents on the perception of metrical rhythm

ILLC DS-2016-04: **Johannes Marti**
Interpreting Linguistic Behavior with Possible World Models

ILLC DS-2016-05: **Phong Lê**
Learning Vector Representations for Sentences - The Recursive Deep Learning Approach

ILLC DS-2016-06: **Gideon Maillette de Buy Wenniger**
Aligning the Foundations of Hierarchical Statistical Machine Translation

ILLC DS-2016-07: **Andreas van Cranenburgh**
Rich Statistical Parsing and Literary Language

ILLC DS-2016-08: **Florian Speelman**
Position-based Quantum Cryptography and Catalytic Computation

ILLC DS-2016-09: **Teresa Piovesan**
Quantum entanglement: insights via graph parameters and conic optimization

ILLC DS-2016-10: **Paula Henk**
Nonstandard Provability for Peano Arithmetic. A Modal Perspective

ILLC DS-2017-01: **Paolo Galeazzi**
Play Without Regret

ILLC DS-2017-02: **Riccardo Pinosio**
The Logic of Kant's Temporal Continuum

ILLC DS-2017-03: **Matthijs Westera**
Exhaustivity and intonation: a unified theory

ILLC DS-2017-04: **Giovanni Cinà**
Categories for the working modal logician

ILLC DS-2017-05: **Shane Noah Steinert-Threlkeld**
Communication and Computation: New Questions About Compositionality

ILLC DS-2017-06: **Peter Hawke**
The Problem of Epistemic Relevance

ILLC DS-2017-07: **Aybüke Özgün**
Evidence in Epistemic Logic: A Topological Perspective

ILLC DS-2017-08: **Raquel Garrido Alhama**
Computational Modelling of Artificial Language Learning: Retention, Recognition & Recurrence

ILLC DS-2017-09: **Miloš Stanojević**
Permutation Forests for Modeling Word Order in Machine Translation

ILLC DS-2018-01: **Berit Janssen**
Retained or Lost in Transmission? Analyzing and Predicting Stability in Dutch Folk Songs

ILLC DS-2018-02: **Hugo Huurdeman**
Supporting the Complex Dynamics of the Information Seeking Process

ILLC DS-2018-03: **Corina Koolen**
Reading beyond the female: The relationship between perception of author gender and literary quality

ILLC DS-2018-04: **Jelle Bruineberg**
Anticipating Affordances: Intentionality in self-organizing brain-body-environment systems

ILLC DS-2018-05: **Joachim Daiber**
Typologically Robust Statistical Machine Translation: Understanding and Exploiting Differences and Similarities Between Languages in Machine Translation