

TOWARDS A DYNAMIC PRAGMATICS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Sven Lauer
August 2013

Abstract

This dissertation explores the interplay of conventional and interactional factors in the interpretation of natural language utterances. It develops a formal framework, dynamic pragmatics, in which pragmatic inferences arise as contextual entailments in a dynamic system in which information states are updated with information about the occurrence of utterance events (in contrast to dynamic semantics, where information states are updated with the content of linguistic expressions). In this way, the framework is able to faithfully model Gricean pragmatic inference as interlocutors' reasoning about each other's utterance choices.

Linguistic utterances are analyzed as having essential effects of two distinct types: Epistemic effects (i.e., effects on the information states of the interlocutors) and normative effects (i.e., effects on the interlocutors' commitments). The latter effects are carried by extra-compositional, normative conventions of use that mediate the form–force mapping; the former arise largely due to the interlocutors' presumptions about each other's beliefs, preferences, and method of determining which (utterance) actions are best (i.e., practical reasoning).

The framework of dynamic pragmatics allows us to consistently take a thoroughly Gricean perspective on language use, and allows us to explore how the interpretation of an utterance arises through the interplay of sentential force, content, and context. At the same time, the framework of dynamic pragmatics sheds a new light on the nature of conversational implicature, and language use in general.

Acknowledgements

There is nothing I can do about it: This paragraph is going to fall short. There is no way to express how much I owe Cleo Condoravdi, and how grateful I am to her, for being my teacher, advisor, mentor, and collaborator. Whenever one leaves a place, there are people and things one leaves behind that one rather would not. I have left many places, but I feel I have never left behind anything so valuable as I will when I leave Stanford and will no longer be able to work closely with Cleo. I will miss her guidance, her understanding, her generosity, her insight . . . my words fail me, so I have to say it simply. Thank you, Cleo.

While Cleo's influence on this dissertation cannot be overstated, Chris Potts also shaped it in many crucial ways. I am indebted to him for his help, advice, encouragement, guidance, enthusiasm and expertise, all of which he has always given freely and generously. He has taught me a lot, and his help was essential in making this dissertation, and its completion, a possibility. And I am grateful to Paul Kiparsky for many valuable comments, for his useful skepticism, and for many helpful discussions. I feel very lucky to have had Cleo, Chris and Paul as my dissertation committee.

I am also grateful to my former teachers, who have all influenced my thinking about the topics of this dissertation: Paul Dekker, Robert van Rooij, Jeroen Groenendijk, Frank Veltman, Peter Bosch, Carla Umbach and Graham Katz.

The ideas that form the basis for this dissertation have gestated in my mind for a long time, and so there are many colleagues and friends who have influenced them over the years. It would be futile to try and compile a complete list, but a few of them deserve special mention. In the following pages, Michael Franke will

recognize many themes familiar from long after-yoga dinner conversations during our joint time in Amsterdam. In many ways, I first started to really think about imperatives and clause types when I sat in Magdalena Kaufmann's ESLLI course and when reading her dissertation thereafter. Her work was a crucial inspiration for Cleo's and my work on these topics. When Anna Chernilovskaya stopped by in Stanford, she was busy finding answers to questions about exclamation marks that I had not even thought to ask, and our collaboration was both fruitful and stimulating. And finally, there is Alex Djalali, who has listened to me ranting and ranted back, who has been my office mate and fellow sufferer, and who I will always regard as my friend and brother. I hereby grant him permission to try and make me uncomfortable by hugging me whenever he pleases.

The Stanford linguistics department has been a wonderful place to write this dissertation, and I am thankful to everyone there. Alex is not the only colleague who became a dear friend. There are also Seung Kyung Kim, Laura Whitton, Jessica Spencer and Natalia Silveira. Getting to know these people alone was worth the struggle of graduate school, the degree is an extra bonus.

This dissertation, and my time at Stanford, would not be the same without the stimulating environment of the rest of the semantics crowd, broadly construed, both former and current, permanent and visiting (in no particular order): Lauri Karttunen, Annie Zaenen, Eve Clark, Ivan Sag, Stanley Peters, Scott Grimm, Nola Stephens, Marie-Catherine de Marneffe, Chigusa Kurumada, Marisa Tice, Eric Acton, Tania Rojas-Esponda, Dasha Popova, James Collins, Lelia Glass, Bonnie Krejci, Tyler Schnoebelen, Mason Chua, Adrian Brasoveanu, Uli Sauerland, Patricia Amaral, Theres Grueter, Dan Lassiter, Tine Breban, Lucas Champollion, Fabio del Prete, and everyone else who ever attended a Meaningful Lunch. This list deliberately leaves out one person, because she needs to be set apart. I cannot thank Beth Levin enough for all her help and for all that she has taught me.

Finally, there is my family. I would not have made it to Stanford, or anywhere, without them: My uncle Dietmar, my aunt Sylvia and the rest of the Holeins. My grandparents Hubert and Elli. My mother Hanne, and my brothers Jan and Lars, who always have my back. I love you very, very much.

Contents

Abstract	iv
Acknowledgements	v
1 Introduction	1
1.1 Going beyond conversational implicatures	1
1.2 The centrality of clause typing	4
1.3 Two kinds of Griceanism	5
1.4 Utterance choice and dynamic pragmatics	7
1.5 Overview of the dissertation	8
2 The basic system	11
2.1 Dynamic pragmatics: The very idea	11
2.2 Communicating contents	14
2.3 The basic system	15
2.3.1 Languages	16
2.3.2 Models for time and belief	17
2.3.3 Events	20
2.3.4 Constraining belief change	23
2.3.5 The dynamic perspective	26
2.4 Communicating contents in the basic system	29
2.5 Communicating with and without intentions	31
2.6 Conclusion	32

3	Clause types	35
3.1	Introducing clause types	35
3.2	Denotation type is not a sufficient guide to function	39
3.3	A case for extra-compositional constraints on use	42
3.4	Illocutionary acts	45
3.5	Sincerity conditions	49
3.6	Summary	51
4	The sentential force of declaratives	53
4.1	Mapping the territory	58
4.2	Declaratives as governed by Lewis-conventions	60
4.3	Declaratives as expressing beliefs	64
4.3.1	Bach and Harnish's (1979) on communicative speech acts . . .	67
4.3.2	R-intentions à la Bach & Harnish	68
4.3.3	Expressing a belief as R-intending	73
4.3.4	Summary	77
4.4	Counts-as rules	77
4.5	Normative theories of clause typing	79
4.5.1	Normative preconditions on utterances	80
4.5.2	Normative effects of utterances	83
4.5.3	Commitment to be held responsible for consequences	87
4.5.4	Commitments to act as though one has a belief	89
4.6	Commitments to act according to an attitude	90
4.6.1	Declaratives with incompatible contents	92
4.6.2	Declaratives express beliefs	92
4.6.3	Other normative consequences of declaratives	93
4.7	The enduring commitments of loose talk: a case for the commitment- to-belief view	95
4.7.1	Loose talk—the basic facts	97
4.7.2	Loose talk in multi-sentence discourses	99
4.7.3	Loose talk in the commitment-to-belief account	101

4.7.4	Communicative reasons for retraction	102
5	Action Choice and Commitment	105
5.1	Modeling action choice	107
5.1.1	Actions and agents	109
5.1.2	Beliefs about action choices	110
5.2	Modeling Preferences	112
5.2.1	Preference structures	113
5.2.2	A working definition for Opt	117
5.2.3	Reasoning about preferences	120
5.2.4	Summary	121
5.3	Modeling commitment	122
5.3.1	Commitments exclude possible future states of the world . .	124
5.3.2	Commitments to beliefs and preferences	127
5.4	Modeling conventions of use	130
5.5	Communicating contents, again	131
5.6	Conclusion	133
6	Commitments to preferences	136
6.1	Imperatives	137
6.1.1	Imperatives as creating commitments to preferences	139
6.1.2	Deriving directive uses	143
6.1.3	Deriving advice uses	148
6.1.4	Conclusion	155
6.2	Performative uses of desideratives	156
6.3	Interrogatives	162
6.4	Denotation types and the form–force mapping	164
6.4.1	Denotations for imperatives	165
6.4.2	Denotations for interrogatives	168
6.4.3	The (possible) indeterminacy of denotation types	170
6.4.4	One convention for declaratives and imperatives?	172
6.4.5	The form of the clause-typing conventions	173

6.4.6	Conclusion	176
6.5	Representing force in the compositional system	177
7	Explicit performatives	181
7.1	Saying makes it so	182
7.2	Explicit performatives as ‘self-verifying’	184
7.3	Truth conditions for performative verbs	187
7.4	Deriving self-verification	189
7.4.1	Introducing illocutionary verbs into <i>Sen</i>	190
7.4.2	Performatives with promise and order are self-verifying . . .	191
7.4.3	Performatives with claim are self-verifying	192
7.4.4	Summary	193
7.5	Further predictions	194
7.5.1	Reportative uses of performative predicates	194
7.5.2	Performatives and logical operators	195
7.5.3	‘Illocutionary’ vs. ‘perlocutionary’ verbs	199
7.6	Comparison to existing approaches	201
7.6.1	Searle (1989): Explicit performatives as declarations	202
7.6.2	Bach and Harnish (1979, 1992)	207
7.7	Conclusion	212
8	Exclamatives and expressives	213
8.1	Exclamatives as an expressive clause type	215
8.1.1	The two implications of exclamatives	215
8.1.2	The nature of the expressive implication	217
8.2	More expressive meanings	224
8.2.1	Non-sentential expressions	224
8.2.2	Non-at-issue meanings	224
8.2.3	Expressive vs. prescriptive conventions	228
8.2.4	Determining the correct convention type	229
8.3	Conclusion	231

9	Conversational implicatures	232
9.1	Preliminaries	235
9.1.1	Maxims and preferences	235
9.1.2	Alternative utterances	237
9.1.3	Two types of preferences and a visual representation of Opt	241
9.2	Some classical implicatures	243
9.2.1	A ‘relevance’ implicature	245
9.2.2	A scalar implicature	247
9.2.3	The ‘epistemic step’	250
9.2.4	Intended implicatures	252
9.2.5	Unintended implicatures	254
9.3	Need a Reason: Mandatory Gricean implicatures	256
9.3.1	Optionality and cancelability	257
9.3.2	The ignorance implicature of disjunction	262
9.3.3	The NaR implicature of disjunction in dynamic pragmatics	266
9.3.4	Generalizing NaR	271
9.3.5	More NaR implicatures?	274
10	Outlook	281
10.1	The role of intentions in pragmatic theory	281
10.2	Ambiguity and underspecification	283
10.3	A question of commitment	285
10.4	Belief revision, salience, and awareness	286
10.5	Conclusion	288
A	The basic system	289
	Bibliography	293

List of Figures

2.1	A branching-time model (from Condoravdi 2002)	18
9.1	Decision in worlds v_{p+rel} where the speaker believes p and takes it to be ‘relevant’	247
9.2	Decision in worlds $v_{p+\neg rel}$ where the speaker believes p and does not take it to be ‘relevant’	247
9.3	Decision in worlds $v_{\Box p}$ where the speaker believes p	248
9.4	Decision in worlds $v_{\Box e}$ where the speaker believes e , but not p	248
9.5	Decision in worlds $v_{\neg\Box e}$ where the speaker believes neither e nor p	248
9.6	Decision in worlds $v_{\Box p+\neg rel}$ where the speaker believes e and p , on the assumption that p is not relevant	249
9.7	Decision in worlds $v_{\Box_i e \wedge \neg p}$ where the speaker knows that e is true and p is false, and believes the implicature will not be drawn	253
9.8	Decision in worlds $v_{\Box_i \neg p \wedge e}$ where the speaker knows that e is true and p is false, and believes the implicature will be drawn	253
9.9	Decision in worlds $v_{\neg\Box p \wedge \neg\Box q}$ where the speaker knows neither p nor q	268
9.10	Decision in worlds $v_{\Box p \wedge \neg\Box q}$ where the speaker knows p but not q	268
9.11	Decision in worlds $v_{\neg\Box p \wedge \Box q}$ where the speaker knows q but not p	268
9.12	Decision in worlds $v_{\Box p \wedge \Box q}$ where the speaker knows both p and q	268
9.13	Decision in worlds where the speaker believes p and has a preference against revealing that	270
9.14	Decision in worlds where the speaker believes p and does not have a preference against revealing that	270

Chapter 1

Introduction

This dissertation is about pragmatics, in a broadly Gricean sense. It is not a dissertation about conversational implicatures, at least not in the classical sense of the term. Implicatures will play a role (Chapter 9 is dedicated to the them), but they are not the central topic. A large part of the dissertation will focus on a question that may seem quite un-Gricean, due to its focus on *linguistic convention*: What kind of linguistic convention makes it so that sentences of different types—such as declaratives, interrogatives and imperatives—are used in different ways, and support different kinds of pragmatic inferences?

This question, however, will be addressed from a very Gricean angle. Pragmatic inference is construed as language users' reasoning about utterance events. Or, more precisely, as language users' reasoning about how utterance events are *chosen*. A central aim of this dissertation is to show that consistently taking such a perspective is fruitful, indeed, necessary if we want to understand language use. The dissertation develops a formal framework, DYNAMIC PRAGMATICS, that enables us to consistently take such a Gricean perspective.

1.1 Going beyond conversational implicatures

Of course, in the theory of Grice (1975), conversational implicatures *are* (intended) inferences about utterance choices. But implicatures, at least as they are usually

understood, are only a special case of such inferences. And limiting oneself to the study of implicature excludes a large number of interesting phenomena from consideration.

Firstly, the classical cases of implicatures are almost always *strengthening inferences*.¹ A speaker utters a declarative sentence, and communicates its truth-conditions. Implicatures are *added* to this communicated meaning.

But not all inferences about utterance choice are of this kind. There are also inferences *weaken* the conveyed content. This is what happens in what Lasnik (1999) calls LOOSE TALK, when a speaker utters a sentence like (1.1) in order to convey that Mary arrived around three o'clock.²

(1.1) Mary arrived at three.

In other cases, what is conveyed by an utterance is neither a logical strengthening nor a weakening of the semantic meaning of the sentence. This is what happens if the audience has reason to doubt the speaker's honesty. If such a speaker utters (1.1), the hearer will not come to believe its semantic content, but he will still draw inferences about why the speaker said what he said (in the way that he said it, at the time that he said it).

Secondly, the Gricean theory of implicature, as originally introduced and usually understood, presupposes a theory of how sentences of different types get associated with their force. The theory starts from the idea that a speaker uttered a declarative sentence *in order to convey information*, that is, in order to make the hearer believe that the truth-conditional content of the sentence is true. But the force of a given sentence, of a given type, in any given context, depends in part on the inferences hearers draw about the speaker's utterance choice. Taking a broader perspective on pragmatic reasoning allows us to model such cases.

Thirdly, in construing pragmatic theory as the study of conversational implicature, somewhat paradoxically, one runs a considerable risk of misunderstanding

¹There is one notable exception, viz., instances where the quality maxim is flouted, as in **And I'm a monkey's uncle**.

²Loose talk will play a crucial role in Chapter 4, in particular Section 4.7.

conversational implicature itself. That is because it is tempting to think of implicatures not as instance of reasoning about language use, but rather to view them as just another kind of implication an utterance may have. That is, it is easy to slip into moving from the perspective articulated in (1.2) into the perspective articulated in (1.3), or even the one articulated in (1.4).

(1.2) *The 'inside' perspective*

Grice taught us that is useful and necessary to think of language use as a species of purposive human behavior. Implicatures are inferences that arise due to interlocutors being aware of this. Their properties follow from the way we derive them.

(1.3) *The 'outside' perspective*

There is a process/module that generates implicatures. Grice gave a theory of how this process/module works and taught us that they have certain properties (optionality, cancelability, non-detachability . . .).

(1.4) *The 'just another implication' perspective*

Conversational implicatures are implications just like at-issue entailments, presuppositions, conventional implicatures etc., which happen to have certain properties, like optionality, cancelability, non-detachability . . .

I don't mean to deny that the 'outside perspective' is useful. It is a convenient abbreviated way to think of pragmatic inference, and it allows us to set aside the details of pragmatic theory when we are working on other issues.

However, I maintain that if we are taking the 'outside' perspective, we have to keep in mind that we have taken a conceptual shortcut, and that the inferences we take for granted arise in a complex manner from factors that influence language use in general.

And I think it is hardly ever prudent to take the 'just another implication' perspective. This perspective obscures the difference between semantic facts (i.e., facts about the grammar of the language) and pragmatic facts (i.e., facts about language use). This is problematic because the reasoning giving rise to pragmatic

sentence types mainly by studying their embedded occurrences. Only in the last decade or so has there been renewed, sustained interest in understanding the uses of various clause types in the formal semantics literature.³ Developing a systematic framework for studying the conventionally specified use of sentences of different types is thus a very timely project.

Secondly, it is this conventional connection between sentences and their use that connects *semantic content*, as it is studied in linguistics, with language use, and inferences about utterance choice. As I just pointed out, the classical Gricean account of implicatures starts from the assumption that a speaker utters a (declarative) sentence in order to convey information, and thus, as I argue in Chapter 3, presupposes an answer to the question of how sentences are conventionally associated with a certain use. And if we want to take a Gricean perspective more generally—if we want to investigate how interlocutors reason about each other’s action choices—we need to know how the contents we study in linguistic semantics relate to the use of sentences. An understanding of clause typing thus is central to developing a formal framework that lets us take a pragmatic perspective in general.

1.3 Two kinds of Griceanism

I have said that this dissertation takes a Gricean perspective on language use. This is correct insofar as it construes pragmatic inference as interlocutors’ reasoning about each other’s utterance choices.

In another way, however, this dissertation takes a rather un-Gricean perspective. It does not involve any central appeal to the notion of *intention* and it does not construe communication and pragmatic inference as being essentially intention-recognition. This is un-Gricean, as Grice famously developed an account of speaker meaning that analyzed the concept in terms of intention (Grice 1957, Grice 1969, Grice 1982). To mean something, for Grice, was to have an intention; to understand

³e.g., Portner (2005, 2007), Schwager (2006), Davis (2009, 2011) Kaufmann (2012), Condoravdi and Lauer (2012) on imperatives; Zanuttini and Portner (2003), Rett (2008, 2011), Castroviejo Miró (2008) on exclamatives; Gunlogson (2003, 2008), Groenendijk and Roelofsen (2009), Davis (2009, 2011), Farkas and Roelofsen (forthcoming) on declaratives and interrogatives.

what someone means was to recognize the corresponding intention.

Now, I do not doubt that when Grice (1975) presented his theory of conversational implicature, and made clear that, for him, a speaker implicates something if he *means* it (but does not *say* it), he had in mind just this kind of speaker-meaning-as-intention. But it seems to me that the basic idea of his account of pragmatic reasoning as reasoning about utterance choices is quite independent from the idea that intentions and their recognition are central to language use.

I do not want to make a grand claim that intentions are irrelevant to understanding language use (and, undoubtedly, in certain circumstances they are), but I will not start from the assumption that intentions are central. Throughout this dissertation, I will occasionally raise the question whether anything crucial has been left unsaid because I have not appealed to intentions, and the conclusion will generally be that it has not. This does not establish the negative claim that intentions are not central to pragmatic reasoning, but it raises the question whether intentions and intention recognition is central to the kind of questions this dissertation aims to answer.

Griceans of a more linguistic bent (e.g., Levinson (2000)) are sometimes accused by Gricean philosophers (e.g., Bach (2012)) of confusing epistemological issues with ontological or metaphysical ones. For example, they are accused of confusing the question of how an addressee can infer the existence of an intended inference with the issue of how it is determined (in the metaphysical sense) whether there is an intended inference. I am sure this charge is sometimes warranted, but I am not certain it is *always* the linguists who are confused. Instead, in many cases, it seems to me that linguistic pragmaticists simply talk about, and are mainly interested in, the 'epistemological' issue of how inferences are derived on the part of the interlocutors and have nothing to say about metaphysical or ontological issues. If a philosopher reads such a linguist's work on the assumption that the linguist is talking about ontological or metaphysical issues, it is not surprising that he will think the linguist is confused.

In large part, this dissertation takes the 'epistemological' perspective, at least as far as pragmatic inference is concerned. I am interested in how interlocutors reason

about each other's utterances, what they learn from each other's utterances, and in how far what they learn is based on their *linguistic* knowledge as opposed to their knowledge about how other people generally behave. This is likely part of the reason why intentions do not seem to play such a great role in the questions I am investigating, while philosophers like Grice and Bach take them to be so central. They are investigating questions different from the ones I am concerned with here.

1.4 Utterance choice and dynamic pragmatics

The framework of DYNAMIC PRAGMATICS developed in this thesis aims to faithfully treat pragmatic inference as interlocutors' reasoning about utterance choice. Utterance choice is construed as an instance of action choice in general. As such, this dissertation is of a piece with recent game-theoretical approaches to pragmatics (Parikh 2001, van Rooij 2004, Benz and van Rooij 2007, Jäger 2007, Franke 2009, Jäger and Ebert 2009, Franke 2011, Jäger 2012, a.o.). It differs from these approaches in two ways. Firstly, it largely abstracts away from the question what the correct 'decision procedure' is that we should assume agents are using when deciding which (utterance) action to perform, while these approaches generally make very specific assumptions about the decision procedure. Secondly, in this dissertation, I will mainly be concerned with the question what *conventional* constraints on use there are, while game-theoretic treatments usually ignore this question.⁴ I see the framework developed here as largely *complementary* to these game-theoretic approaches, rather than as an alternative to them.

Besides being strongly inspired, as the game-theoretic treatments are, by Grice's work, the framework developed here also owes much to the work of Stalnaker (1978, et seq.) and that of other authors building on his insights. Stalnaker's conception has sometimes been referred to as 'dynamic pragmatics' (e.g., by Schlenker (2010, p. 390)), though I am not aware that Stalnaker has used the label himself. As

⁴Most game-theoretic models of pragmatics employ a semantic interpretation function that is assumed to be given by convention, but they do not invoke any conventional constraints on use proper.

in Stalnaker's work, a crucial role is played by interlocutors *public beliefs* (though I construe this notion somewhat differently, cf. Chapter 4) and how they get updated in the course of a conversation. The conception offered here differs in that it assumes that there also are *public preferences*, and in that it emphasizes the *normative character* of both notions. Finally, it explicitly models *action choice*, and reasoning about action choice, which doesn't usually play a big role in Stalnaker's writings on language.⁵

1.5 Overview of the dissertation

The following chapters develop a formal framework for pragmatic reasoning that integrates an articulated theory of clause typing and applies the framework to a number of phenomena. The first five chapters alternate between introducing the formal set-up and conceptual and empirical considerations that motivate it, while the chapters thereafter apply the framework to specific phenomena. The progression is as follows.

Chapter 2 introduces the basic idea of a dynamic pragmatics as the term is understood here, and sets up the basics of the formal framework. The exposition uses as its running example a very simple pragmatic inference, viz., the inference to the truth of a declarative utterance.

Chapter 3 introduces some of the basic questions raised by the existence of different clause types. It argues that the association between sentences of different types and their uses must be conventional in nature, and lays out some basic assumptions about clause typing that are made in this dissertation.

Chapter 4 focuses on declarative sentences. It explores various hypotheses about what their conventionally-specified use is, and what kind of convention specifies it. The chapter concludes with an informal characterization of my own

⁵Of course, it does play a big role in his writings on the epistemic foundations of game-theory (Stalnaker 1994, Stalnaker 1996), which also have shaped, though more indirectly, some of the ideas in this dissertation.

proposal, building on work in Condoravdi and Lauer (2011, 2012) and Lauer (2012). Utterances of declaratives are argued to commit their speaker to a belief in the truth of the uttered sentence.

Chapter 5 extends the framework introduced in Chapter 2 with the tools necessary to formally implement the theory of declaratives proposed in Chapter 4. Action choice, preferences and commitment. The chapter concludes with reconstructing the account of the pragmatic inference in Chapter 2.

Chapter 6 moves beyond declarative sentences. The main focus is on *imperatives*, which are claimed to commit their speaker to a *preference* instead of a *belief*. The main focus of the chapter is to demonstrate that the framework of dynamic pragmatics allows us to show how the varied uses of imperatives arise in context from an interaction of semantic content, sentential force, and interactional reasoning.

Chapter 7 takes up the issue of *explicitly performative utterances*, which have played a great role in speech act theory. It integrates the analysis of Condoravdi and Lauer (2011) into the current framework, and shows how the central properties of these sentences arise straightforwardly from an interaction of the proposed lexical meanings with the pragmatic system.

Chapter 8 offers some preliminary considerations on the question whether *all* conventional constraints on use should be understood in terms of commitments, as the previous chapters have proposed for declaratives, imperatives and interrogatives. I argue that this is not the case, and that we should recognize different kinds of conventions, of the kind proposed by Lewis (1969), in particular for *exclamative sentences* and various ‘expressive’ items.

Chapter 9 deals with conversational implicatures. It shows how some standard cases can be treated in the framework of dynamic pragmatics, mainly to illustrate how various ‘optimization-based’ theories of implicatures (including the game-theoretic ones mentioned above) fit into the current set-up. The chapter

then goes on to show how the framework of dynamic pragmatics lets us appreciate a significant, and surprising, prediction that such optimization-based theories (including Grice's own) make. Gricean pragmatic inferences can be *mandatory*, i.e., neither cancelable nor optional. This has significant consequences, both because optionality and cancelability have often been used as tests for implicature-hood, and because it potentially extends the domain of Gricean pragmatics in a significant way. Phenomena that usually have been taken to be outside the reach of Gricean explanations may be amenable to a Gricean treatment after all.

Chapter 2

The basic system

In this chapter, I will work towards a model of a very simple pragmatic inference—so simple, indeed, that at first glance it may not seem like a pragmatic inference at all. Suppose *Ad* is on the phone with *Sp*, and *Sp* utters (2.1).

(2.1) It is raining in Chicago.

From observing *Sp*'s utterance in (2.1), *Ad* can learn, if the context is right, that it is raining in Chicago. This is the inference this chapter is dedicated to modeling. In doing so, I will set up the foundations of the framework of dynamic pragmatics that will be developed and applied in the rest of this dissertation.

2.1 Dynamic pragmatics: The very idea

The general idea of a DYNAMIC PRAGMATICS is to appropriate some of the tools of DYNAMIC SEMANTICS to model information change. The basic idea of dynamic semantics is that utterances change *information states*, which are taken to either represent what the addressee of an utterance knows, what the speaker knows, what is COMMON GROUND between the interlocutors, or something of the kind. The various incarnations of dynamic semantics (Kamp 1981, Heim 1983, Groenendijk and Stokhof 1991, Groenendijk, Stokhof and Veltman 1995, Veltman 1996, Beaver

2001, among many others) thus start by specifying a formal representation of such information states (sets of possible worlds, files, discourse representation structures, assignment functions, etc.), and then analyze the meaning of linguistic expression as UPDATE POTENTIALS—that is, update operations on information states (usually represented as functions or relations over information states). We can schematize the idea as in (2.2):

$$(2.2) \quad \text{new info state} = \\ \text{old info state} + \text{meaning of the uttered expression}$$

For present purposes, we can identify the meaning of an expression with its *informational content*, and so we can instantiate this general schema for the sentence in (2.1) as in (2.3).

$$(2.3) \quad \text{new info state} = \\ \text{old info state} + \text{the information that it is raining in Chicago}$$

In dynamic pragmatics, we also conceive of utterances as changing information states, but the information we add is of a different kind: It is the information *that a certain utterance has happened*. The information states that we update are intended to model the beliefs of the addressee (or more generally, any interlocutor). Schematically, for (2.1):

$$(2.4) \quad \text{new info state} = \\ \text{old info state} + \text{the information that } Sp \text{ uttered } \mathbf{It\ is\ raining\ in\ Chicago.}$$

I immediately introduce two idealizing assumptions. Firstly, I will assume throughout that utterance observation is perfect: When a speaker utters a sentence, his interlocutors will perceive it, and not doubt their perceptual capacities, i.e., they will come to believe that the utterance happened. Secondly, I abstract away from all issues of ambiguity and semantic underspecification, and essentially assume that interlocutors observe utterances of disambiguated logical forms.

Both of these assumptions are, of course, rarely warranted in everyday discourse, but making them will keep our framework simpler. In the concluding Chapter 10, I will return to this assumption and indicate how it could be lifted. In the meantime, we will see that there is plenty of complexity to deal with even if we ignore the possibility that hearers mishear, or have false beliefs about, or are uncertain about, the value of contextual parameters.

Information states will be modeled, as in Veltman (1996)'s update semantics, as sets of possible worlds. That is, the information state of agent *Ad* is the set of possible worlds that are compatible with what *Ad* believes. Then, we can model information gain as the *removal* of worlds from an information state. Learning that *p* is modeled by removing all those worlds from the information state that do not make *p* true. With this, our schema becomes:

$$(2.5) \quad \text{new info state} = \\ \text{old info state} \cap \{w \mid Sp \text{ uttered } \mathbf{It\ is\ raining\ in\ Chicago\ in\ } w\}$$

Given this general possible-worlds setup, it makes sense to identify the informational content of **It is raining in Chicago** with a set of possible worlds, as well. Then we can contrast the dynamic pragmatics schema in (2.5) with the dynamic semantic schema in (2.6) (this is just the update that happens in Veltman (1996)'s system for non-modal sentences):

$$(2.6) \quad \text{new info state} = \\ \text{old info state} \cap \{w \mid \text{It is raining in Chicago in } w\}$$

We can think of (2.5) as a 'cautious update' and (2.6) as a 'credulous update': With (2.6), we model something that happens if an addressee observes an utterance of **It is raining in Chicago** and believes it to be true. With (2.5), by contrast, we model what happens *whenever* an addressee observes an utterance of **It is raining in Chicago**: The addressee automatically adds to his information state is what he can be certain about, *viz.*, that the utterance he just observed happened.

Much of this chapter is dedicated to further spelling this out formally. In order to do so, we must specify, for example, what it is for an utterance to occur at a world. Before doing so (in Section 2.3), I want to briefly preview how to derive the inference to the truth of the content.

2.2 Communicating contents

We ultimately want to model how it is that a speaker comes to believe in the truth of the content of an observed utterance. One way to do this would be to simply assume that *two* updates happen: The one in (2.5) and the one in (2.6). This is the idea in much work building the conception of assertion in Stalnaker (1974, 1978, 1994, 2002): When an utterance like **It is raining in Chicago** happens, the update in (2.5) is performed automatically, and at the same time, the speaker *proposes* to also perform the update in (2.6).

This is not what is intended here: The *only* update that happens when a speaker makes an utterance is the one in (2.5). But this update, will, if the context is right, also inform the addressee that the uttered sentence is true. That is, in certain contexts, the content of an utterance will be a *contextual entailment* of the fact that the utterance was made. In which contexts? The typical case is one in which the addressee takes the speaker to be both honest and well-informed with respect to the topic of his utterance.

If the addressee takes the speaker to be honest, (2.7) will be true in all worlds in his information state (where \Rightarrow is material implication):

(2.7) $(Sp \text{ utters } \mathbf{It\ is\ raining\ in\ Chicago}) \Rightarrow (Sp \text{ believes it is raining in Chicago})$

And if the addressee takes the speaker to be well-informed, all worlds in his information state will also make true statement in (2.8):

(2.8) $(S \text{ believes it is raining in Chicago}) \Rightarrow (\text{It is raining in Chicago})$

But if both (2.7) and (2.8) is true in a given world, so is (2.9):

(2.9) (S utters **It is raining in Chicago**) \Rightarrow (It is raining in Chicago)

This means that the worlds in the addressee's information state fall into one of two classes:

1. Worlds in which the speaker utters **It is raining in Chicago** and it is raining in Chicago.
2. Worlds in which the speaker does not utter **It is raining in Chicago**.

The addressee observes the speaker's utterance of **It is raining in Chicago**. That is, we remove all worlds from the information state in which the speaker has not made this utterance.

But that means that we are left with an information state that only contains worlds in the first class—and these are all worlds in which it is raining in Chicago. That is, in virtue of his prior belief that the speaker is honest and knowledgeable about rain in Chicago, learning that the speaker uttered **It is raining in Chicago** also means, automatically, learning that it is raining in Chicago. If the addressee's pre-utterance belief state is of the right kind, then performing the update in (2.5) will also eliminate all worlds in which it is not raining in Chicago.

2.3 The basic system

In the present section, I will set up the basic ingredients for framework of dynamic pragmatics. I will only introduce the ingredients necessary to model the simple inference to the truth of the content that I just summarized informally. Chapter 5 will extend this basic system into one in which more pragmatic phenomena can be treated.

2.3.1 Languages

The object language $Prop$

To keep things maximally simple, the system I set up here will model pragmatic inference in a community of speakers of classical propositional logic. This has the advantage that we will not have to worry whether the semantics given for the language is adequate to any particular group of speakers, as, by assumption, our fictional speech community uses the language with its classical semantics. In Chapter 3 and Chapter 7, I will very minimally extend the object language, but its basic propositional character will remain the same.

The formal definition of the language $Prop$ and its semantics is given in Appendix A. The semantics is an entirely standard intensional one: Proposition letters are interpreted as a set of possible worlds, conjunction is interpreted as set intersection, negation as set complement, and so forth.

The pragmatic language \mathcal{P}_{Prop}

Given an object language L , we define a corresponding PRAGMATIC LANGUAGE \mathcal{P}_L . The pragmatic language \mathcal{P}_{Prop} , by necessity, much richer than our $Prop$ itself. We want to be able to talk about *agents* (speakers, addressees) and other individuals, and their *beliefs*. Further, we want to be able to talk about *time* and we want to be able to quantify over individuals and times.

To distinguish formulas of the object language and the pragmatic language, I will use indexed versions of φ as meta-variables for object language formulas, and indexed versions of ϕ and ψ as meta-variables for formulas of the pragmatic language.

I will use i, i', i'' etc. to refer to *agents* (interlocutors) and t, t' etc. to refer to times. The pragmatic language is a multimodal language whose modalities are indexed to individuals and times. Thus, to represent the fact that i believes ϕ at time t , we write:

$$(2.10) \quad \Box_{i,t}\phi$$

‘At time t , agent i believes that ϕ .’

\mathcal{P}_{Prop} also shares its propositional constants with $Prop$, and it has the same propositional connectives. As a result, $Prop$ is a sublanguage of \mathcal{P}_{Prop} , and we can write things like:

$$(2.11) \quad \Box_{i,t}(\neg p \wedge q)$$

‘At t , i believes that p is false and q is true.’

And we have the usual first-order quantification over individuals and times, so we can represent:

$$(2.12) \quad \begin{array}{l} \text{a. } \exists_x : \Box_{x,t}p \\ \text{b. } \exists_t : \Box_{i,t}p \\ \text{c. } \exists_x : \Box_{x,t}\exists_y : \Box_{y,t}p \end{array}$$

‘There is an agent that believes, at t , that p .’

‘There is a time at which i believes that p .’

‘There is an agent that believes that there is another agent who believes that p .’

2.3.2 Models for time and belief

In Section 2.2, we conceived of information states as sets of possible worlds. In order to talk about beliefs and time, we will conceive of the worlds in an information state as being drawn from a *background model* that has some additional structure. I will introduce the background models first and then, in Section 2.3.5, show how we can define dynamic belief states in terms of them. The result will be that we take a dynamic *perspective* on the static background models.

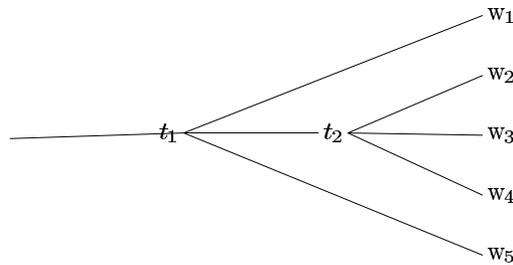


Figure 2.1: A branching-time model (from Condoravdi 2002)

Forward-branching time

The formulas of \mathcal{P}_{prop} are interpreted on ‘forward-branching’ models of the $T \times W$ -type (Thomason 1984). Models like this have been fruitfully employed in the linguistics literature to capture the interaction of tense, modality and conditionals by Condoravdi (2002), Kaufmann (2005) and Kaufmann and Schwager (2009), among others. Forward-branching models capture the intuition that the past of a world is fixed, but its future is not. At every given time, there are multiple ways the world could develop in the future. The structure of the models can be visualized as in Fig. 2.1.

This forward-branching structure is constructed in the following way. The set T of times is linearly ordered. Worlds are taken to determine complete courses of affairs at all times. Frames determine a time-indexed equivalence relation \approx_t between worlds. $w_1 \approx_t w_2$ intuitively means ‘ w_1 and w_2 share their history at least up to time t ’. For a given world w , the equivalence class $\{v \in W \mid w \approx_t v\}$ can be thought of the set of *possible futures of w at time t* —that is, at t , all worlds in this set are ‘the same’, the worlds in the equivalence class constitute all the ways how the world could evolve after t .

To obtain the frame in Fig. 2.1, we let

$$(2.13) \quad \begin{aligned} \text{a.} \quad & W = \{w_1, w_2, w_3, w_4, w_5\} \\ \text{b.} \quad & T = \{t_1, t_2, t_3\} \\ \text{c.} \quad & t_1 < t_2 < t_3 \\ \text{d.} \quad & w_1 \approx_{t_1} w_2 \approx_{t_1} w_3 \approx_{t_1} w_4 \approx_{t_1} w_5 \\ & w_2 \approx_{t_2} w_3 \approx_{t_2} w_4 \\ & \approx_{t_3} = \emptyset \end{aligned}$$

Thomason's definition of $T \times W$ -frames is in Definition 4 in Appendix A. Throughout this dissertation, I will assume that time is discrete, and in fact that T is just the set of natural numbers with its usual ordering. This is sufficient for present purposes and has the advantage that we can refer to 'the time just after t ' as $t + 1$. I do not assume, however, that all branches have a common root—i.e., that \approx_0 is total, as Fig. 2.1 might suggest.

Beliefs

In order to interpret the belief-modalities $\Box_{i,t}$, we extend the basic $T \times W$ -models with accessibility relations of the kind familiar from the relational models introduced by Kanger (1957), Hintikka (1961) and Kripke (1963). For given i, t , $R_{i,t}$ is a relation on W that is transitive, Euclidean and serial.¹ Modal operators are interpreted using this relation in the usual way, resulting in an *SD45*-logic for $\Box_{i,t}$. In particular this ensures the usual introspection properties:

$$(2.14) \quad \begin{aligned} \text{a.} \quad & \text{POSITIVE INTROSPECTION} \\ & w \vDash \Box_{i,t}(\phi) \Rightarrow w \vDash \Box_{i,t}(\Box_{i,t}\phi) \\ \text{b.} \quad & \text{NEGATIVE INTROSPECTION} \\ & w \vDash \neg\Box_{i,t}(\phi) \Rightarrow w \vDash \Box_{i,t}(\neg\Box_{i,t}\phi) \end{aligned}$$

¹A relation on a set X is *transitive* if xRy and yRz implies xRz , it is *Euclidean* if xRy and xRz imply yRz and it is *serial* for all $x \in X$ there is a y such that xRy .

Variables and constants are interpreted rigidly, i.e., independently from the world. In order to ensure that beliefs respect the structure of the branching-time model, we impose two constraints:²

$$(2.15) \quad \text{HISTORICITY constraint}$$

$$\text{If } w_1 \approx_t w_2, \text{ then } w_1 R_{i,t} v \text{ iff } w_2 R_{i,t} v$$

$$(2.16) \quad \text{NO FORE-BELIEF constraint}$$

$$\text{If } v_1 \approx_t v_2, \text{ then } w R_{i,t} v_1 \text{ iff } w R_{i,t} v_2$$

(2.15) requires that the belief relations respect \approx , i.e., if two worlds have not divided at t , an agent must have the same beliefs in both worlds. (2.16) captures the idea that, if two worlds are undivided at t , then it is ‘objectively unsettled’ which one of the two will become actual—in that case, an agent should either take both worlds to be possible, or neither. This is a very intuitive constraint as long as we think of the objective uncertainty to be due to ‘external factors’. In Chapter 5, we will see that we have to weaken it in the context of action choice.

We also introduce an operator for common belief at a time C_t , which is interpreted using the transitive closure of the individual belief relations at t :

$$(2.17) \quad C_t \phi$$

$$\text{‘At } t, \text{ it is common belief that } \phi\text{.’}$$

As usual, this means that ϕ is common belief iff everyone believes that ϕ , everyone believes that everyone believes that ϕ , everyone believes that everyone that everyone believes that ϕ , and so on.

2.3.3 Events

The final ingredient in the basic system is the notion of an *event*. We introduce these in two steps. First, we define a set of *event types*, which are model-theoretic

²The name ‘historicity’ for (2.15) is taken from Kaufmann (2005). He calls (2.16) ‘No fore-knowledge’. The full constraints on accessibility relations are summarized in Definition 9 in Appendix A.

entities standing for classes of events with fixed participants. Then, we define a function that specifies which event type is instantiated in a given world, at a given time, and introduce *event predicates* into \mathcal{P}_{Prop} that allow us to talk about events in the language.

Event types

I assume a stock of *event classes* \mathbb{E} . Throughout this dissertation, we will be largely using only one such class, *utter*. Event types are composed of such an event class, together with a set of arguments (participants). *utter* takes three arguments, two individuals and a formula of the object language, forming event types such as:

$$(2.18) \quad \text{utter}(i_1, i_2, \varphi)$$

‘an event of i_1 uttering φ towards i_2 ’

Note that even though these event types have the form of atomic predications, they are *not* formulas of \mathcal{P}_{Prop} . Instead, they are model-theoretic entities. The individual ‘arguments’ are not constants or variables referring to entities, but the entities themselves. This means that the formulas of $Prop$ are also part of the models for \mathcal{P}_{Prop} , about which I will have more to say below.

Event instantiation

Models for \mathcal{P}_{Prop} contain a partial function Hap (for ‘happens’), that specifies which event types are instantiated at a world and time. For a given event type $E(a_1, \dots, a_n)$, $\text{Hap}_w(t_1, t_2) = E(a_1, \dots, a_n)$ intuitively stands for ‘ $E(a_1, \dots, a_n)$ happens between t_1 and t_2 ’.

Assuming Hap is a function is a useful simplification. To simplify even further, I restrict attention to models in which $\text{Hap}_w(t, t')$ is only defined if $t' = t + 1$ —this means that events only happen ‘between’ two adjacent time-steps (i.e., events are instantaneous), and there is at most one event happening at any given time.³

³These assumptions that could be lifted, but they will make the subsequent definitions run much more smoothly. Of course, in a general setting, we want to allow for events to overlap temporally,

Further, I impose the following two more constraints on **Hap**. The first is a minimal requirement, ensuring that **Hap** respects the temporal structure of our models, the second is a useful simplification.

(2.19) HISTORICITY OF **Hap** constraint

If $w_1 \approx_t w_2$ then for all $t_1, t_2 \leq t$: $\mathbf{Hap}_{w_1}(t_1, t_2) = \mathbf{Hap}_{w_2}(t_1, t_2)$

(2.20) *Determinism* constraint

If $w_1 \approx_t w_2$ and $\mathbf{Hap}_{w_1}(t, t+1) = \mathbf{Hap}_{w_2}(t, t+1)$ then $w_1 \approx_{t+1} w_2$.

(2.19) ensures that if an event type is instantiated in one world, but not the other, the two worlds diverge. (2.20) requires the converse, and is arguably conceptually dubious.

According to (2.20), worlds *only* divide if different event types are instantiated in them—which means that the type of an event fully determines the effect the event has in a given world. This is likely too strong (if not for conceptual reasons, then because the event types that we will be using are too ‘coarse’), but in the present context, the assumption does no harm, and it makes it a little easier to intuitively grasp the shape of our models. Given an initial specification of \approx_0 (for the first time step 0), \approx is fully determined by **Hap**.

Event predicates

We have introduced *event types* in our models, as well as a function that specifies when an event happens. But so far, we cannot yet talk about events in the language \mathcal{P}_{Prop} . In order to do so, we assume the language contains *event predicate letters*, which are standard predicate letters, indexed with event constants or variables.

For what is to follow, we again need only one such predicate letter, **utter**. Instead of interpreting it via an interpretation function that is then suitably constrained to give **utter** its intended interpretation, I state the satisfaction conditions of **utter**-formulas directly:

and allow for non-instantaneous events so as to reason about what happens while an event is underway, but neither will be necessary for what is to follow.

$$(2.21) \quad w \models \text{utter}_e(a, b, \ulcorner \varphi \urcorner) \text{ iff } \text{Hap}_w(I(e)) = \text{utter}(I(a), I(b), \varphi)$$

Note that one of the arguments of `utter` contains a formula of *Prop*. Implicitly, this definition treats $\ulcorner \cdot \urcorner$ as an operator that takes a *Prop*-formula (which, as indicated above, are part of our models) into a *name* that refers to this formula.⁴ Much like with parentheses, I will often omit the $\ulcorner \cdot \urcorner$ operator from example formulas, unless it aids readability.

Note that the symbol `utter` is overloaded—it figures both in event types and in event predications. Similarly, it will frequently be useful to sloppily use a single symbol to refer to an individual in a model and a constant of the language that refers to this individual—e.g., I will often use *Sp* to refer to the speaker-as-agent in the model, but also use *Sp* as a constant that implicitly is assumed to refer to the model-theoretic entity *Sp*. I trust no confusion will arise from this.

Finally, I define two more versions of `utter` which take fewer arguments. The other arguments are existentially closed: $\text{utter}_e(i, \ulcorner \varphi \urcorner)$ is true if *e* is an utterance of φ by *i* towards an arbitrary agent, $\text{utter}_e(i)$ is true if *e* is an utterances by *i* of an arbitrary sentence towards an arbitrary agent.

2.3.4 Constraining belief change

So far, I have not said anything substantial about how beliefs evolve over time. I specified some general constraints requiring that the basic structure of the underlying frame is respected, but nothing prevents an agent from having wildly different beliefs at adjacent times *t* and *t'*. We remedy this in the present section.

Ensuring utterance uptake

It is here where we encode, in terms of a constraint on admissible models, the idealizing assumption that agents perfectly observe each other's utterances.

⁴Potts (2005, Appendix A.3) makes essentially the same move—his logic \mathcal{L}_U , which in many ways can be seen as similar to \mathcal{P} also contains constants that refer to sentences of his 'object' language \mathcal{L}_{CI} .

(2.22) PAL CONSTRAINT

For all i, w, w', t :

if there is v such that $wR_{i,t}v$ and $\text{Hap}_w(t, t + 1) = \text{Hap}_v(t, t + 1)$, then

$$wR_{i,t+1}w' \text{ iff } wR_{i,t}w' \text{ and } \text{Hap}_w(t, t + 1) = \text{Hap}_{w'}(t, t + 1)$$

This formulation is quite complex, because we have to take into account that an agent can have false beliefs, and hence can observe events that he took to be impossible. If we excluded that possibility (or did not require the belief relations to be serial), we could impose the simpler constraint in (2.23):

$$(2.23) \quad R_{i,t+1} = \{ \langle w, v \rangle \in R_{i,t} \mid \text{Hap}_w(t, t + 1) = \text{Hap}_v(t, t + 1) \}$$

What (2.23) says is that when an agent observes an event, this eliminates all links from his belief relation that point to worlds in which the event does not happen. (2.22) imposes the same requirement, except that it leaves unconstrained what happens if an agent observes an event he antecedently believed would not occur. In full generality, we would want to impose additional constraints to cover this case, which ensure that the necessary *revision* of beliefs is minimal. How to achieve this is a difficult topic, and it is orthogonal to the concerns in this dissertation, and so I leave belief revision unconstrained, except for assuming that the result of revision ensures that the agent believes that the event he just observed happened.

(2.22) does two things: (i) it ensures that beliefs only change in light of events⁵ and (ii) it ensures that all agents have perfect knowledge about all events that occur. Given that this is globally the case throughout our models, whenever an event happens, it will thereafter be *common knowledge* that it did, and events happening is the only way in which knowledge changes. As a consequence, all acquired knowledge is shared knowledge. This is surely inadequate in general—people can and do learn things privately, and pass on this information later, and that needs to ultimately be reflected in a theory of pragmatics.

⁵With one exception: If an agent was certain that an event would occur in the next time step, but none occurs, this will trigger belief revision.

However, for present purposes, this conception will be sufficient. Recall that the events we are mainly interested in are *utterance events* happening in the course of a conversation. If we assume that all agents are party to the conversation, assuming that they have common knowledge of all utterance events is not so crazy. Essentially, the idealizing assumption we are making with (2.22) is that all events are local to the interlocutors, that all interlocutors (correctly) observe all events and that they also *observe each other* observing the events.

The connection to Public Announcement Logic (PAL)

Modeling information gain as the removal of links from an accessibility relation is not a novel idea at all. It goes back at least to Landman (1986). Semantics for PUBLIC ANNOUNCEMENT LOGIC (PAL)⁶ often make use of the idea. Indeed, putting the constraint in (2.22) on our models means that we can construe our timelines (modulo belief revision) as *sequences of PAL models* where the model at each time step is derived from the previous one by the ‘public announcement’ that the event happened.

In PAL, updates can be made with arbitrary formulas, including those that contain knowledge (or belief) operators.⁷ In the present system, ‘updates’ only happen with *event formulas*. And the conceptualization is quite different. In PAL, ‘announcements’ are classically construed as utterances that provide the information that their content is true. This makes PAL quite limited in modeling conversation in general, as it essentially presupposes that all utterances are honest and based on perfect information.

But in our present system, we do not have ‘public announcements’ of the *content* of utterances, but rather *of the fact that an utterance happened*—that is we update with $\text{utter}(a, \varphi)$, not with φ . Instead of ‘public announcement’, we should instead

⁶More precisely: Those semantics that are not stated as a special case of a more general framework, such as Dynamic Epistemic Logic (DEL), that is, cf. van Ditmarsch, van der Hoek and Kooi (2008).

⁷That is why in PAL, it is not generally ensured that ϕ is commonly known after a public announcement, i.e., $[\phi]C_i\phi$ is not a theorem of PAL—it is invalidated in cases like $\phi = p \wedge \neg\Box p$ (‘You don’t know yet that p ’)—but we disallow such updates, as we update only with *event expressions*, not with arbitrary formulas.

perhaps speak of ‘joint observation’. But the logic remains essentially the same.

Such a conception of PAL is not without precedence in the literature:

“So far, all announcements were made by an agent that was also modelled in the system. We can also imagine an announcement as a ‘public event’ that does not involve an agent. Such an event publicly ‘reveals’ the truth of the announced formula. Therefore, announcements have also been called ‘revelations’—announcements by the divine agent, that are obviously true without questioning.”

(van Ditmarsch et al. 2008, p.76)

Except that in the present case, where ‘announcements’ only concern publicly-observable events, we need not invoke a divine announcer—we can replace him by the agent’s ability to observe (and observe each other observing). The tight relation between the way beliefs change in the current model and PAL suggests that generalizations of PAL are useful resources for lifting the simplifying assumptions about perfect uptake made here. I return to this issue in Chapter 10.

In summary, the PAL constraint ensures that, if an event happens, all agents learn that it did, and this is the only information they acquire. Given that this is common knowledge in our models, that also means that when an event happens, it becomes common knowledge that it did.

2.3.5 The dynamic perspective

So far, the system I have introduced is not quite *dynamic*. It is a model of how beliefs change, to be sure, but our models themselves are entirely static. The overall $T \times W$ model is fixed once and for all. It takes “the view ‘from above’, viewing epistemic temporal models as a Grand Stage where events unfold” (van Benthem and Pacuit 2006, p. 88).

By contrast, the ‘preview’ I gave in Section 2.2 was phrased in entirely dynamic terms—taking, in van Benthem and Pacuit’s terms, the ‘view from below’. In the

present section, I introduce a way to take a ‘local’ perspective on the ‘global’ Grand Stage model.

The Grand Stage model is useful to set up the system, to define constraints regulating how beliefs should evolve, and to specify which non-epistemic consequences events have on the world (which we will do in Chapter 5). But the local, ‘dynamic’ perspective is easier to handle if we want to talk about what happens at a particular time step. This is why, in the later parts of the thesis, when we apply the model developed here and in Chapter 5, we will almost exclusively take the dynamic perspective. For a fixed \mathcal{P}_{Prop} -model, we define:

$$(2.24) \quad B_{i,t,w} := \{v \in W \mid wR_{i,t}v\}$$

‘the belief state of i at t, w ’

$B_{i,t,w}$ simply collects the worlds that are compatible with what i believes (in world w , at time t). For such belief states, we then define the update operation $[\cdot]$:

(2.25) Let ev be an event formula, then

$$B_{i,t,w}[ev] := \{v \in W \mid v \in B_{i,t+1,w} \text{ and } \text{Hap}_v(t, t+1) = ev\}$$

‘the belief state that i would be in if event ev happened just after t ’

If the (immediate) occurrence of ev is not ruled out⁸ in $B_{i,t,w}$, this is obviously equivalent (given the PAL constraint above) to:

$$(2.26) \quad B_{i,t,w}[ev] = \{v \in B_{i,t,w} \mid \text{Hap}_v(t, t+1) = ev\}$$

(2.26) is essentially the update familiar from Veltman (1996)’s *update semantics*, except that we don’t update with formulas of the language, but with event type formulas. We define a notion of *support* for information states, as follows:⁹

⁸In case ev was believed to be impossible, the right-hand side of (2.26) is empty, but the same is not true for $B_{i,t,w}[ev]$, which in this case is derived by belief revision.

⁹Veltman ultimately defines support in terms of vacuous update:

$$(i) \quad B \vDash \phi \text{ iff } B[\phi] = B$$

(2.27) For any information state B and \mathcal{P}_{Prop} formula ϕ :
 $B \models \phi$ iff for all $v \in B : v \models \phi$

With this, we obviously have:

(2.28) $B_{i,t,w} \models \phi$ iff $w \models \Box_{i,t}\phi$

We can thus think of the dynamic perspective as ‘zooming in’ to (an agent’s perspective at) a given world and time in the $T \times W$ model. The benefit of having a background model is that we immediately obtain results like the following:

(2.29) For any two agents i, i' , we have, at all t, w :
 $B_{i,t,w}[\text{utter}(a, b, \varphi)] \models \Box_{i',t+1} \text{utter}_t(a, b, \varphi)$

And the stronger:

(2.30) $B_{i,t,w}[\text{utter}(a, b, \varphi)] \models C_{t+1} \text{utter}_t(a, b, \varphi)$

The benefit of the dynamic perspective, by contrast, is that it is much easier to handle, and more perspicuous, than the sometimes unwieldy background model. To state (2.29) in terms of the background model, for example, we would have to say instead:

(2.31) For all i, i', t, w : If $\text{Hap}_w(t, t+1) = \text{utter}(a, b, \varphi)$, then
 $w \models \Box_{i,t+1}\Box_{i',t+1} \text{utter}_t(a, b, \varphi)$

(2.29) makes much easier to see, at a glance, what is going on. We thus can think of the dynamic perspective as an *interface* to the global model. But it is also more than just an interface, as it stresses the dynamic nature of what we are modeling.

In many cases when I use the dynamic perspective in the rest of this dissertation, I will actually suppress all temporal parameters (and, when appropriate, also the world parameters). Often, these will be irrelevant because throughout this thesis,

We do not have this option here, as we do not update with formulas, but with event expressions. Luckily, since all formulas can be evaluated *pointwise*, we lose nothing by defining support in the distributive manner done in (2.27).

I will be concerned with one-shot utterances. The basic set-up of the model has obvious potential to be extended so as to allow reasoning about *discourse planning*, involving multiple acts in succession, but we will not deal with such things here. Even when the value of the temporal indexes are relevant, there intended values will often be clear from the context. This suggests that, ultimately, we could simplify the dynamic perspective further, by having it use a simplified version of the full pragmatic language without time parameters. However, for the space of this thesis, I shall simply leave these parameters implicit wherever possible. If we do so, then (2.29) becomes the very readable (2.32):

$$(2.32) \quad \text{For any two agents } i, i', \text{ we have:}$$

$$B_i[\text{utter}(a, b, \varphi)] \vDash \Box_{i'}(\text{utter}(a, b, \varphi))$$

2.4 Communicating contents in the basic system

We have just seen a very general result about updates:

$$(2.33) \quad B_{i,t,w}[\text{utter}(a, b, \varphi)] \vDash C_{t+1} \text{utter}_t(a, b, \varphi)$$

(2.33) holds for all i, t, w in all models for \mathcal{P}_{Prop} . Actually, this is about the strongest thing we can say about the general case. But we can now make *contextual assumptions* that allow us to derive more interesting things. In particular, we can straightforwardly capture the way in which an addressee comes to believe in the truth of what a speaker says, if he takes the speaker to be well-informed and honest (I drop the world parameter in what follows).

$$(2.34) \quad \text{Contextual assumptions: Trusting addressee}$$

a. 'Honest speaker'

$$B_{Ad,t} \vDash \text{utter}_t(Sp, \varphi) \Rightarrow \Box_{Sp,t}(\varphi)$$

b. 'Informed speaker'

$$B_{Ad,t} \vDash \Box_{Sp,t}(\varphi) \Rightarrow \varphi$$

If (2.34a) is true, then so is (2.35):

$$(2.35) \quad B_{Ad,t}[\text{utter}(Sp, Ad, \varphi)] \models \Box_{Sp,t}(\varphi)$$

And if, in addition, (2.34b) is true, so is (2.36):

$$(2.36) \quad B_{Ad,t}[\text{utter}(Sp, \varphi)] \models \varphi$$

So we capture the fact that if a speaker is taken to be honest and knowledgeable, then learning that he uttered φ amounts to learning that φ is true. Surely, this is not a great achievement of our system, but we now have a very general system in place for reasoning about utterance events and their (epistemic) consequences. That it validates such pervasive inference is reassuring.

Actually, the system gives us a little more. Suppose the conditions in (2.35) are believed to be true by the *speaker* about the addressee's belief state.¹⁰ Then we also have:

$$(2.37) \quad B_{Sp,t}[\text{utter}(Sp, \varphi)] \models \Box_{Ad,t+1}\varphi$$

That is, if a speaker thinks his addressee takes him to be honest and knowledgeable, he will also believe that if he utters φ , his addressee will come to believe φ , too. That is, we can model the fact that the speaker is aware of the *epistemic consequences* of his potential actions.

In Chapter 5, we will see how this gives us a way to reason about utterance choice. Suppose a speaker has to decide whether to utter p , q , $p \vee q$ or $p \wedge q$ —then he has to compare the expected consequences of doing each. That is, he has to compare the information states in (2.38) with respect to how well they satisfy his goals.

¹⁰That is, we have

- (i) a. $B_{Sp,t} \models \Box_{Ad,t}(\text{utter}_t(Sp, \varphi) \Rightarrow \Box_{Sp,t}\varphi)$
- b. $B_{Sp,t} \models \Box_{Ad}((\Box_{Sp,t}\varphi) \Rightarrow \varphi)$

- (2.38) a. $B_{Sp,t}[\text{utter}(Sp,p)]$
 b. $B_{Sp,t}[\text{utter}(Sp,q)]$
 c. $B_{Sp,t}[\text{utter}(Sp,p \vee q)]$
 d. $B_{Sp,t}[\text{utter}(Sp,p \wedge q)]$

In order to do this, of course, we need to be able to talk about goals and their satisfaction. That is what Chapter 5 will be all about.

2.5 Communicating with and without intentions

An interesting fact about the way we just derived (2.36) is that we did not make any references to the speaker's *intentions* or *desires*. In particular, we did not assume that the speaker intended or desired to communicate anything—we just assumed that the addressee believed that the speaker is honest and well-informed.

One might question, then, whether what I have described above is truly communication—it is an instance of *information transfer* (at least if we assume that the speaker indeed believes what he said), but for communication we might want to require that the speaker also *intends* that the addressee form the belief in question.¹¹

Even this is not always taken to be sufficient for true communication (or at least successful communication) to occur. Bach and Harnish (1979) maintain that communication is successful only if the addressee also *recognizes* the speaker's intention to make him believe something.¹² But on a certain level of description, there is little difference between situations in which an addressee comes to correctly recognize the intention of the speaker to communicate something, and those in

¹¹Though English **communicate** at least has one sense on which this is not required, as attested by examples like (i).

(i) In some false confession cases, details of the crime are inadvertently communicated to a suspect by police during questioning.
 (<http://www.innocenceproject.org/fix/False-Confessions.php>, last retrieved on August 21, 2013)

¹²In fact, this is more the view of Grice (1957)—I am not sure what exactly the intention is that Bach and Harnish (1979) claim the hearer must recognize for communication to succeed—see Section 4.3.2 for discussion.

which he otherwise believes that the speaker would only utter sentences he believes to be true.

Of course, this assumption *may* be justified, on a particular occasion, by the following two beliefs:

- (2.39) a. The speaker will utter φ only if he intends me to believe it.
 b. The speaker would not intend me to believe anything he does not believe himself.

(2.39a) is true and believed by the addressee, then observing an utterance will lead the addressee to recognize that the speaker intends to communicate φ —but even this is not necessarily enough for a communicative intention à la Bach and Harnish to succeed—for, they say, a communicative intention is also ‘intended to be recognized as intended to be recognized’. So it is not enough if an addressee recognizes the speaker’s intention to make him believe something—he must also recognize that the speaker intended him to recognize the intention.

At this point we start to wonder whether all this intention recognition is relevant in everyday linguistic practice. Suppose a speaker wants to pass on a certain piece of information—say, that it is raining in Chicago—to someone else. He utters a sentence—say, **It is raining in Chicago**—and the addressee takes him to be honest and well-informed, and so he comes to believe that it is raining in Chicago. Does it matter, to the speaker or the addressee, *why* the addressee was predisposed to accept the speaker’s utterance as truth? Has communication failed, in an interesting sense, if the addressee fails to also believe that the speaker intended him to recognize that he intended him to recognize that he intended him to believe that it is raining in Chicago?

2.6 Conclusion

In this chapter, I have introduced a number of the foundational notions necessary for the framework of dynamic pragmatics: a formal system that allows us to talk

about beliefs and how they develop over time and about the occurrence of events and how they influence agents' beliefs. A particular subset of events, *utterance events* are of particular importance, as one would expect from a system that is intended for linguistic pragmatics. I have then illustrated the basic workings of the system by showing how we can model a pragmatic inference that other systems—in particular many systems of dynamic semantics—take for granted: I have shown how the model accounts for the fact that an addressee who takes the speaker to be honest and knowledgeable can come to believe in the informational content of an utterance he observes.

Before moving on, I want to briefly stop and reflect on a particular feature of the system as set up so far. The information that gets added to our agents' belief states in the course of a conversation is, in some sense, not particularly linguistic in nature. It is simply the fact that an event happened. On a very basic level, utterance events are on a par with non-linguistic and non-communicative events. Thus, to use Stalnaker's oft-cited example, basic gain in information by observing an utterance event is on a par with the basic gain in information by observing that a goat just walked in.

Let us dwell on this for a second. Our event types could include, for example, an event *donkey_walks_in(d)*¹³ and if we introduce the corresponding predicates, whose interpretation is suitably specified, we could derive facts as such as the following:

$$(2.40) \quad B_{i,t,w}[donkey_walks_in(d)] \models \exists x : donkey(x)$$

'If *i* observes a donkey walking in, he comes to believe that a donkey exists.'

This may give one pause. Is the system of dynamic pragmatics supposed to model *every* aspect of our cognitive experience of the world? Does a full specification of the model require an account of every kind of cognitive effect of an observation? This seems crazy, and it would be.

¹³This is a particularly weird representation for such an event, of course—but this is part of my point here: The current system, as a system of dynamic pragmatics, is not intended to give plausible representations of arbitrary events. It is interested in giving plausible representations of *utterance events*.

On a very abstract level, the system of dynamic pragmatics indeed can be seen as a model of *any* kind of belief formation in response to events. I think this is appropriate—it embodies the very Gricean assumption that pragmatic reasoning is continuous, at least in principle, with other kinds of reasoning.

But all that means is that the system of dynamic pragmatics, as set up so far, *could*, in principle, be extended to model non-linguistic reasoning. A non-linguist *could* appropriate the model, adding his own event types and predicates and suitably constraining their interpretation to model reasoning about non-linguistic events. But in doing so, he would have to extend the system, making assumptions that are appropriate to the domain he is modeling. Take the example in (2.40): In assuming that there is an event type such as *donkey_walks_in(d)* (which gets perfectly observed), we have to assume that our agents have the concept of a donkey, and that they can (perfectly) distinguish donkeys from non-donkeys, that their concept of donkeys is compatible with donkeys changing location, that they have a concept of ‘coming in’, and so forth. These assumptions may be reasonable (as idealizations) for certain purposes, but they are completely *non-linguistic* assumptions, which is why I have not made them.¹⁴

The assumptions that I have made instead—in assuming that there is an event type *utter*, and that there is a corresponding event predicate, and the way it relates to the event type—are *thoroughly linguistic*. I have assumed that there are such things as utterances of sentences, and that agents recognize them as such. And the idealization I have made in assuming that agents observe utterances of disambiguated logical forms amounts to the assumption of (idealized) linguistic competence: I assume not only that agents recognize utterances as such, but also that they have the appropriate competence to analyze the phonological, morphological, syntactic and semantic structure of the utterance so as to assign truth conditions to it.

Much of the rest of this dissertation will be concerned with the question of what *other* linguistic knowledge we need to assume our agents to have beyond this.

¹⁴Of course, a linguist might sometimes make some of these assumptions, for example when studying the semantics of prepositions like *in*.

On some level of description, all the sentences in (3.1) involve the same *content*, which can be characterized as the proposition that is true iff the addressee will come tomorrow. However, utterances of these sentences employ this proposition in very different ways: An utterance of (3.1a) intuitively *claims* that this proposition is true, while an utterance of (3.1b) *inquires* whether the proposition is true and an utterance of (3.1c) attempts to *induce* the addressee to make the proposition true. A central question for a pragmatic theory to answer is how this difference comes about, and more generally how content, sentence type and context interact in language use.

To this end, we define a minimal extension of the object language *Prop*, as in (3.2).

$$(3.2) \quad \text{Sen} := \mathcal{L}_+ \cup \mathcal{L}_? \cup \mathcal{L}_!, \text{ where}$$

- a. $\mathcal{L}_+ := \{ \vdash \phi \mid \phi \in \text{Prop} \}$
- b. $\mathcal{L}_? := \{ ?\phi \mid \phi \in \text{Prop} \}$
- c. $\mathcal{L}_! := \{ !\phi \mid \phi \in \text{Prop} \}$

The question at hand now is how utterances of sentences in \mathcal{L}_+ differ from utterances of sentences in $\mathcal{L}_?$, and how both differ from utterances of sentences in $\mathcal{L}_!$. The answer to this question should tell us why sentences in \mathcal{L}_+ are typically used for informing or claiming (and in which contexts they can be used for other purposes), while sentences in $\mathcal{L}_?$ are typically used to request information (and in which contexts they can be used for other purposes), while sentences in $\mathcal{L}_!$ are typically used in an attempt to get the addressee to do something (and in which contexts they can be used for other purposes).

At the end of Chapter 2, when showing how information can be conveyed by means of a utterance (which was implicitly taken to be the utterance of a declarative), I used the assumption that the addressee believes that the speaker would only utter a declarative ϕ if he believed ϕ to be true. The question about \mathcal{L}_+ we are facing now is how to justify this assumption.

That this assumption needs justification, even in cases of perfect cooperation, may not be obvious: Isn't it simply more cooperative to speak the truth than to

speak falsity? Grice (1975) certainly thought so: His MAXIM OF QUALITY purportedly is motivated by his COOPERATIVE PRINCIPLE:

(3.3) MAXIM OF QUALITY (Grice 1975, p. 46)

Try to make your contribution one that is true.

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

(3.4) COOPERATIVE PRINCIPLE (Grice 1975, p. 45)

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

Of course (3.3) does not follow from (3.4) in any direct fashion. And it should not, for (3.3) only is intuitively valid for declaratives, but not for interrogatives and imperatives. So, the maxim should really be formulated as in (3.5).

(3.5) MAXIM OF QUALITY FOR DECLARATIVES

Try to utter only those declaratives that are true.

1. Do not utter declaratives that you believe to be false.
2. Do not utter declaratives for whose truth you lack adequate evidence.

This new version makes it more obvious that the maxim should not follow from assumptions about cooperative behavior alone. Why should general considerations about what constitutes cooperative behavior, by themselves, say anything about sentences of a syntactically individuated class of sentences? There must be some linguistic property of declaratives, that, together with general considerations of cooperative behavior, allows us to derive a maxim like (3.5) (in contexts where the maxim is appropriate).

Something that plausibly *could* be derived from general considerations about cooperative behavior is a maxim like the one in (3.6).²

²A formulation of the quality maxim that is quite close to (3.6) has been used by Joshi (1982) in the context of question answering to account for clarification follow-ups.

(3.6) MAXIM OF QUALITY FOR COMMUNICATED CONTENTS

Do not intentionally make your addressee believe anything that is not true.

1. Do not intentionally make your addressee believe something that you believe to be false.
2. Do not intentionally make your addressee believe something for which you lack adequate evidence.

If we want to derive (3.5) from (3.6), we need to explain why declaratives generally have the effect of making the addressee believe their contents—but that is just what we wanted the QUALITY maxim to help us explain!

Put differently: Grice's QUALITY maxim only sounds like a plausible principle of cooperative behavior if we assume that declaratives usually have the effect (or are usually intended to have the effect) of making the addressee believe their content—but then, this maxim cannot be the reason for this being the typical function of declaratives.

We could derive similar maxims for interrogatives and imperatives—if we have some independent means of deriving their (typical) functions:

(3.7) MAXIM OF 'QUALITY' FOR INTERROGATIVES³

Try to utter only those interrogatives that you want your addressee to answer truthfully.

(3.8) MAXIM OF 'QUALITY' FOR IMPERATIVES

Try to utter only those imperatives that you want your addressee to fulfill.

But again, as maxims of conversation, these principles are not the *explanans* that tells us why interrogatives and imperatives have the uses they do. Rather, they

³If we were to adopt such a maxim, we might also want to add the following sub-maxim:

- (i) Try to utter only those interrogatives that you don't know the answer to.

Though ((i)) is not appropriate for all utterances of interrogatives—*rhetorical questions* and *exam questions* are counter-examples. Similarly (3.7) arguably is not obeyed by certain kinds of combatively-used interrogatives; see Section 6.3.

are an *explanandum*: They should follow from general principles of cooperative behavior, together with what is independently known about interrogatives and imperatives.

Clearly, it is not just general considerations about cooperative behavior that allow us to explain why sentences of different types have the uses they have. This is fairly obvious for non-declaratives, but considering the question in the setting of multiple clause types makes us realize that it is true for declaratives just as well.

3.2 Denotation type is not a sufficient guide to function

I have introduced the syntax of *Sen*, but I have not yet specified the semantics of expressions of the language. In formal semantics, it is general practice to assign different kinds of denotations to sentences of different types. For declaratives and interrogatives, this can be motivated on semantic grounds alone, without much reference to pragmatics: Both declaratives and interrogatives can be embedded in larger sentences, and the semantics of the larger sentences puts certain constraints on what the semantics of the embedded sentences can be. So, while it may be suitable to assume that declarative sentences denote intensional propositions (sets of possible worlds), we cannot assume that interrogatives denote such propositions: It is hard to see what proposition a **wh**-question like **who will come to the party** could denote so that we can correctly derive the intuitive truth-conditions of sentences like **John wondered who will come to the party** and **John decided who will come to the party** and **We argued about who will come to the party**.⁴

This difference in logical type, however, is not sufficient to explain the differences in use. *Some* of these difference may be traced to differences in denotation

⁴Recent work in the INQUISITIVE SEMANTICS paradigm (Groenendijk and Roelofsen 2009, Ciardelli and Roelofsen 2011, a. o., building on the work of Hamblin (1958)) has argued that we can (and should) assign a uniform semantic type to declaratives and interrogatives—crucially, this value must be more complex than a flat set of worlds, i.e., more complex than a proposition in the usual intensional sense. An *inquisitive proposition* is a set of sets of worlds, which classically is assumed to be the semantic value only of interrogatives.

type, but not all. For example, one might reason as follows: If interrogatives denote, say, sets of sets of possible worlds, while declaratives denote simply sets of possible worlds, maybe this is enough to explain why sentences are used in the way they are. Indeed, even independently from considerations of formal semantics, there is a strong intuition that the notion of *truth* does not properly apply to interrogatives and imperatives. Doesn't that suffice to explain why the QUALITY maxim applies only to declaratives?

Maybe it does,⁵ but then we only have explained why interrogatives are not (outside of special contexts) used to make claims or provide information; we have not yet explained why declaratives typically are. We might venture on, though, and argue as follows: The denotation type of declaratives simply is something that is uniquely suited to conveying information and making claims, at least in standard contexts. After all, one might think, declaratives (or, rather, their denotations), are something that can be true or false at a world—so *of course* they are used to claim that their contents are true, or to express the belief that their contents are true.

We can illustrate this possible line of thinking by means of an analogy: On a given day, I know that my roommate has visited the German bakery in Palo Alto. Returning home late at night from a session of dissertation writing, I find a loaf of delicious bread sitting on the desk in my room. Considering the situation, what I know about the uses breads can be put to, and about my roommate, I understand that the bread is a gift, and that he intends me to eat and enjoy it.

Similarly, the story would go, when a speaker utters a declarative, he puts in front of the addressee a certain semantic object (say, a function of type $\langle s, t \rangle$), and given what the addressee knows about the purposes that this type of semantic object can be put to, he can reason to the conclusion that the speaker intends to provide him with some factual information.

But this cannot work, at least if we maintain the assumption that matrix sentences have the same denotation as the sentence would have when embedded. The reason is this: Given what we know about possible embedding environments,

⁵Though note that a set of possible worlds, in itself, has no special relation to the notion of *truth*—it acquires such a relation only when paired with a claim or judgement (to use Frege's term) that the actual world is contained in the set.

the semantic value of declaratives (and interrogatives) must be very versatile. Whatever an embedded declarative denotes, it must be something that can be: believed, disbelieved, preferred, dispreferred, surprised about, hypothesized, inferred, doubted, and many other things, given that declaratives can be embedded under predicates such as **believe**, **disbelieve**, **prefer (want)**, **disprefer**, **surprised**, **hypothesize**, **infer**, **doubt** and so forth.⁶ Similar considerations apply to interrogatives: Whatever their denotation type is, it must be something that can be: asked, known, agreed on, found out, pondered, decided, and many other things in virtue of the fact that there is a multitude of question-embedding predicates.

To drive this point home: Assume, for example, that declaratives denote sets of possible worlds. In itself, such a set is no more suited to expressing a belief that the actual world is a member of the set than for many other purposes. For example, such a set is equally well-suited for expressing the belief, or conveying the information, that the actual world is *not* in this set, or that the speaker *desires* the actual world to be in this set, or that he desires it *not* to be in this set, and many other things. And the problem is not that sets of possible worlds are the wrong kind of denotatum: As long as we assume that sentences have the same denotation in matrix and embedded uses, this problem necessarily arises: A semantic object that somehow is uniquely suited for expressing beliefs will not do as the denotation of many embedded declaratives and a semantic object that is suited as a denotation for all embedded declaratives will not be uniquely suited for expressing beliefs.

Thus, while the study of embedded sentences can tell us something about what kind of denotation a sentence should have, the nature of this denotation will not be enough to sufficiently constrain the uses such a sentence can be put to when it is not embedded. We need something extra in order to explain why matrix sentences

⁶I speak here of the denotation being ‘something that can be believed’, etc., which may suggest that this kind of consideration only applies if we assume ‘static’ objects in the semantics, rather than the kind of ‘dynamic’ denotations familiar from dynamic semantics. But this is not so: Even if we take declaratives to denote, say, update relations between information states, in order to account for the semantics of declarative-embedding predicates, we need to assume that these relations hold between agents’ belief states, their desire states, and so forth. So a dynamic denotatum, too, must be very versatile. Too versatile to make providing information or expressing a belief the ‘obvious’ or only thing a speaker could be trying to do with his declarative utterance.

of a particular type are used the way they are.

3.3 A case for extra-compositional constraints on use

So we need something extra—over and above the denotation sentences of various types have when they are embedded—in order to explain the constraints on use that we observe. One way to go would be to put this additional information into the denotation of matrix sentences—thus giving up the idea that matrix sentences have the same denotation as their embedded counterparts.

Perhaps the most straightforward way of doing this is to assume that matrix sentences uniformly contain an operator that scopes over the whole sentence—call it a ‘force’ operator. We then could hypothesize a different such operator for each sentence type. This would amount to the claim that natural languages are in fact structured like *Sen*, with its three operators \vdash , $?$ and $!$, and every matrix sentence is headed by such an operator. The denotational semantics of these operators then would be responsible for specifying, or at least constraining, the possible uses of the sentence.

This kind of analysis has been proposed by Krifka (2001b, to appear), who takes the force operators to denote *speech acts* (in particular, ILLOCUTIONARY ACTS in the sense of Austin (1962), a feature I discuss in Section 3.4).

There is, however, another option, and this is the one I will pursue in the following: We can specify the context changes brought about by an utterance *directly*, as a ‘convention of use’ that is followed by the speech community. On this construal, we can maintain the assumption that matrix sentences have the same denotation as embedded sentences. The constraints on use are captured extra-compositionally, in the conventions of use.

One reason to do things this way is that ultimately, regardless of the denotation one assigns to matrix sentences, be it static or dynamic, the denotation, on its own, cannot solve the basic problem. What is asked for is a link between the denotation on the one hand, and use on the other. Since the denotation is one of the two things to be linked, it cannot be the link itself. If we assume that matrix sentences denote

speech acts, for example, we still will need a convention, rule or principle like (3.9).

(3.9) **Performance principle** (hypothetical)

If a speaker utters a sentence that denotes a speech act *a*, *a* is thereby performed.

Similarly, if sentences denote some kind of update operation on contexts, we still would need a convention, rule or principle like (3.10).

(3.10) **Context-change principle** (hypothetical)

If a speaker utters a sentence that denotes a context change operation, this operation is thereby applied to the context.

In either way, we will have to assume something extra to go from what the sentence *denotes* to the effect of its *use*. But once we grant that something is needed in addition to a compositionally determined denotation—some kind of principle, rule, or convention that is, by necessity, external to the system of semantic composition—this additional component may well be the proper place to represent conventional constraints on use. The principles in (3.9) and (3.10) don't represent any constraint of use by themselves—they just enable compositionally-encoded constraints to have an effect. The alternative is to have multiple principles—one per sentence type—that specify the dynamic effect of the utterance directly. These conventions could make reference to the morpho-syntactic characteristics of sentence types directly (as in (3.11)), or reference a denotation type that is particular to the sentence type (as in (3.12)).⁷

(3.11) a. **Declarative principle** (hypothetical)

If a speaker utters a sentence that is morpho-syntactically declarative, this has the following effect: . . .

b. **Interrogative principle** (hypothetical)

⁷The conception of clause-typing developed in this dissertation—which builds on Condoravdi and Lauer (2011, 2012)—is neutral with respect to which of these two (or a number of other) options is chosen, as I extensively demonstrate in Section 6.4.

If a speaker utters a sentence that is morpho-syntactically interrogative, this has the following effect:

c. . . .

(3.12) a. **Declarative principle** (hypothetical)

If a speaker utters a sentence that has a denotation of type $\langle s, t \rangle$, this has the following effect: . . .

b. **Interrogative principle** (hypothetical)

If a speaker utters a sentence that has a denotation of type $\langle \langle s, t \rangle, t \rangle$, this has the following effect:

c. . . .

At first glance, it may seem that a general principle such as (3.9) is more uniform—but this uniformity comes at a price: We have to complicate the system of semantic composition, we have to give up the assumption that embedded sentences and matrix sentences have the same denotation, and we have to assume that every sentence contains a silent operator that is otherwise unmotivated. Moreover, just as we need one principle per sentence type in (3.12), we would need one silent operator per sentence type in the compositional system—and then, *in addition*, a general principle that relates denotations to use. So, on balance, putting information about force into the compositional system is *less* parsimonious than keeping it outside of the system. At the same time, having extra-compositional principles regulate the dynamic effects of sentences seems intuitively quite appropriate. After all, what we want to model is constraints on possible uses. What better place to put these constraints than in the principle(s) that connect denotation and use?

I shall hence proceed on the assumption that the form–force mapping is mediated by extra-compositional principles or conventions of use. Once the basic concepts are in place, I will return (in Section 6.5) to the question whether a representation of force in the compositional system is necessary.

3.4 Illocutionary acts

I have just argued that, if it can be done, it is more parsimonious and conceptually attractive to represent conventional constraints on use outside of the system of semantic composition. Before we proceed, in the next chapters, to the question of what the nature of these constraints should be, it is useful to set aside a popular conception of modeling the effects of utterances and the constraints on use that derive from these effects, viz., that of speech act theory in the sense of Austin (1962) and Searle (1969, et seq.).

As mentioned above, Krifka (2001b, to appear) takes matrix sentences to denote high-level ILLOCUTIONARY ACTS such as ASSERT and ORDER, etc. By analogy, one could think that the extra-compositional conventions I envision map sentences to the same kind of act:

- (3.13) a. **Declarative illocutionary principle** (hypothetical)
 If a speaker utters a sentence that is morpho-syntactically declarative (or: that has a denotation of type $\langle s, t \rangle$), he thereby ASSERTS the content of the sentence
- b. **Interrogative illocutionary principle** (hypothetical)
 If a speaker utters a sentence that is morpho-syntactically interrogative (or: that has a denotation of type $\langle \langle s, t \rangle, t \rangle$), he thereby QUESTIONS the content of the sentence
- c. **Imperative illocutionary principle** (hypothetical)
 If a speaker utters a sentence that is morpho-syntactically imperative (or: that has a denotation of type X), he thereby COMMANDS the addressee to fulfill the content of the imperative.

It is well-known, however, that this cannot work: Sentences of a given type are not always used to perform the same high-level illocutionary act, even when they are used entirely literally and sincerely. This is particularly easy to illustrate in the case of imperatives: As Schmerling (1982) pointed out, imperatives can be used to perform a wide variety of illocutionary acts (Kaufmann (2012, based on

the author's dissertation, Schwager (2006)) calls this the PROBLEM OF FUNCTIONAL INHOMOGENEITY):

- | | | | |
|--------|----|---|---------------|
| (3.14) | a. | Stand at attention! | (COMMAND) |
| | b. | Don't touch the hot plate! | (WARNING) |
| | c. | Hand me the salt, please. | (REQUEST) |
| | d. | Do the right thing! | (EXHORTATION) |
| | e. | Take these pills for a week. | (ADVICE) |
| | f. | Please, lend me the money! | (PLEA) |
| | g. | Get well soon! | (WELL-WISH) |
| | h. | Drop dead! | (CURSE) |
| | i. | Please, don't rain! | (ABSENT WISH) |
| | j. | Okay, go out and play. | (PERMISSION) |
| | k. | Okay then, sue me, since it's come to that. | (CONCESSION) |
| | l. | Have a cookie(, if you like). | (OFFER) |

A conceptually unattractive reaction to this problem would be to disjunctively list the possible illocutionary forces in the 'imperative principle':

(3.15) **Imperative illocutionary principle** (hypothetical)

If a speaker utters a sentence that is morpho-syntactically imperative (or: that has a denotation of type X), he thereby COMMANDS or REQUESTS or WISHES or ADVISES or ... the addressee to fulfill the content of the imperative.

Similarly (and just as unattractively), in a Krifka-style analysis, we could assume that imperatives (and sentences of all other types) are multiply ambiguous, with a large range of possible force operators (all of them silent) that all can occur on the matrix level in imperatives.

Instead, the approach pursued here is to assume that imperatives (and sentences of other types) have a uniform force on all their (sincere) uses. This force, then, must be quite general. It is only in context that this force gets strengthened to

something as specific as ORDER. Borrowing a term from Chierchia and McConnell-Ginet (1990), we can call the (general, underspecified) force assigned to sentences by linguistic convention their SENTENTIAL FORCE, to distinguish it from the concept of illocutionary force. The theory I will lay out in the following will not make any essential use of the concept of illocutionary force. Nor will it involve a concept that is anything like it, at least to the extent that illocutionary force is understood in terms of high-level acts such as PROMISE, ORDER or ASSERT.

The reason for this is that I do not see any advantage in a two-step procedure that first maps utterances to illocutionary acts and then relates these types to their use in context. Such a two-step procedure would of course be advantageous if either the mapping from sentence types to illocutionary acts or the mapping from illocutionary acts to purposes of interlocutors would be homogeneous (ideally, functional). And it would be almost mandated if both mappings were homogeneous.

But neither mapping is, in fact, homogeneous. We have already seen that imperatives do not generally correspond to a single illocutionary type (cf. (3.14)), and it is no more homogeneous for other sentence types. And illocutionary acts of any type can be performed for all kinds of reasons (in Austin's terms, with all kinds of perlocutionary intentions), and give rise to all kinds of inferences about the speaker's preferences, beliefs and choices. It is hence far from obvious that employing a level of description that uses high-level illocutionary acts as the basic units of organization is helpful.

Further, not only is it dubious that employing a level of illocutionary description is theoretically helpful, it also seems that which exact kind of illocutionary act has been performed is often irrelevant in linguistic practice: Suppose *A* and *B* are in a car, *B* is driving, *A* has a map. The following interaction takes place:

- (3.16) *A*: Take a left here.
 B: [*turns left*]

What is the illocutionary force of *A*'s utterance? Is it an order? A request? A suggestion? All three? Neither? More importantly: Does it matter? *A* wanted *B* to turn left. *B* observed *A*'s utterance, concluded that *A* wanted him to turn left,

knew that their current goals are aligned, and turned left. *B* did not have to decide what the illocutionary force of *A*'s statement is, nor is it clear that *A* intended any particular illocutionary force. He intended to get *B* to turn left, and thought the imperative was a good means to get him to do this, and it was, that is all. It is not clear that we gain anything in describing this instance of communication by making reference to illocutionary acts.⁸

I hence reject the idea (implicit in Searle's work, and quite explicit in authors such as Bach and Harnish (1979)) that what interlocutors do in conversation, first and foremost, is to try and recognize each other's 'illocutionary intentions', i.e., that it is somehow a fundamental part of interpretation to identify which illocutionary force a speaker has intended. Something like this may be crucial in certain, quite special contexts (e.g., when a boss says to his employee **You look awful, take the rest of the week off**—in this case, it *may* be crucial whether this counts as an order or merely a piece of advice), but it is not central to communication in general.

Giving up the idea that it is paramount to explain how utterances of sentences of a given type relate to illocutionary types (and how illocutionary types relate to pragmatic inference) opens up the possibility of a simpler, and quite attractive, understanding of the role sentence types play in language use. Uttering a sentence of a given type has certain consequences, determined by linguistic conventions. These consequences, together with facts about the contents and inferences about language users' goals, explain patterns of language use.

This is not to say that the acts that Austin and Searle singled out as illocutionary acts are somehow suspect and that labels for these acts are not occasionally useful. They are quite helpful as descriptive labels for uses-in-context. The labels on the examples in (3.14) are a perspicuous way to make the point that different imperative utterances have quite different effects in context. Secondly, there are natural language predicates that correspond, at least roughly, to illocutionary types:

⁸Of course, one could retreat and say that *A*'s utterance had some weak or underspecified illocutionary force. But this just is the point: It is not clear that we need anything other than such weak or underspecified forces—forces that are general enough so that we can assume they stand in a one-to-one relation with sentence types. That is, we need *sentential forces*, but it is not clear that we need high-level illocutionary forces of the kind discussed by Searle and Austin.

order, promise, assert, and so forth. Ultimately, we will want to specify the lexical meanings of these verbs, and in so doing, we may well need to appeal to some of the notions that have played a role in philosophical speech act theory. But the fact that there are such verbs, in itself, is no reason to assume that the concepts involved in their lexical semantics play a particularly central role in how communication works.⁹ And the existence of such verbs (which may specify quite particular contextual effects of the utterance) is sufficient to explain why these verbs are highly useful for descriptively categorizing sets of uses-in-context.

There is *one* property of ‘illocutionary verbs’ (i.e., verbs that roughly correspond to the illocutionary acts studied by speech act theorists) that sets them apart: They can be used in ‘explicitly performative sentences’, such as (3.17).

(3.17) I (hereby) promise to come.

In Chapter 7 I will show how the semantics of such verbs can be specified without making reference to corresponding illocutionary acts, in a way that explains the peculiar behavior of sentences like (3.17).

3.5 Sincerity conditions

There is one concept from speech act theory that one might take to be central in our understanding of the form–force mapping, even if one rejects the centrality of illocutionary notions: The concept of sincerity conditions (or ‘felicity conditions’ in general). Indeed, it might appear that talking in terms of sincerity conditions could enable the linguist interested in the semantics of clause types to leave aside any complex considerations about what kind of effect an utterance of a sentence has in context.

One may grant that there are, or seem to be, properly linguistic conventions involved in the form–force mapping, and that linguists should investigate them.

⁹After all, there are also various verbs whose lexical semantics refers to the volume and related acoustic properties of an utterance, such as **whisper, shout**, etc. From this, we do not conclude that ‘volume’ is a central notion for our explanation of language use.

But maybe we just need to find a way, *any* way, to state the content of these conventions, and then move on to studying linguistic facts that are unadulterated by facts about use.

More concretely, could we not just capture the conventional effects of clauses of different types on the model of truth conditions, by stating sincerity conditions for utterances, leaving the concept of sincerity unanalyzed?

We certainly could do that. We could say that a complete description of a language requires the characterization of a function *FORCE* which takes utterance events into the set {*SINCERE*, *INSINCERE*}. For English, a partial characterization of this function could look as follows:

(3.18) *FORCE* is a function such that:

- a. if *u* is an utterance of a declarative, $FORCE(u) = SINCERE$ iff the speaker of *u* believes the content of *u*.
- b. if *u* is an utterance of an interrogative, $FORCE(u) = SINCERE$ iff the speaker of *u* wants the addressee to tell him which of the propositions in the content of *u* is true.
- c. if *u* is an utterance of an imperative, $FORCE(u) = SINCERE$ iff . . .

We would then insist that the job of the linguist is done with the specification of *FORCE*. Again in analogy with truth conditions, we would say that just as it is not the job of a linguist to explain the notion of truth, it is not the job of a linguist to explain the notion of sincerity. All a linguist needs to do is to explicate properly linguistic constraints on sincerity.

Let us grant for a moment that this kind of architecture is the right one, and that the putative clean separation of labor between linguistics and other fields responsible for pragmatics (such as psychology and/or philosophy) is practicable and desirable. Then we still would need an answer, however preliminary and subject to revision, to the question what the conventional effects of utterances of sentences are, and what role these effects play in the inferences speaker-hearers draw about each other's utterance choices. But this will require, in turn, an answer,

however preliminary and subject to revision, to how the concepts *SINCERE* and *INSINCERE* relate to language use; why it is a fact, if it is a fact, that speakers generally strive to make utterances that are *SINCERE*, and under what conditions this generalization does and does not apply.

That is, even if we grant that, in some sense, a theory answering such questions is outside the domain of linguistics proper, linguists still *need* such a theory—if only for methodological reasons. It is well-known, for example, that there is no pre-theoretical way to distinguish conversational implicatures from entailments—but then, without a theory of how pragmatic inference works (and how it works for sentences of all types), it is unclear how the data (actual utterances and native speaker’s intuitions about possible utterances) relates to our linguistic theories.

Perhaps more importantly, the assumption that there is a function like *FORCE* which maps utterances (or utterances and contexts) into $\{SINCERE, INSINCERE\}$ is not as theoretically harmless as it may seem. Sincerity (or felicity) conditions are not just any specification of constraints on use, but rather a quite particular one. Assuming that a specification of sincerity (or felicity) conditions is the correct way to account for clause-typing amounts to the assumptions that what matters are *preconditions*, i.e., rules that say when under what circumstances a sentence can be sincerely uttered. But this is far from the only logical possibility. Indeed, as I will argue in the following chapters, it is not in such preconditions, but rather in their *consequences* (that is, *postconditions*) that utterances of sentences of different types differ.

3.6 Summary

Sentences of different types are used in different ways, and support different kinds of pragmatic inferences. In this chapter, I have pointed out that this cannot be explained either by general principles of cooperative behavior, nor by the fact that they have denotations of different kinds, at least if these denotations are anything like the denotations these sentences have when embedded in other sentences.

I have further argued that it may be possible (and, if possible, desirable) to

capture conventional constraints on use outside of the system of semantic composition, and have cast doubt on the usefulness of the concept of illocutionary acts in understanding language use. The latter two points are preliminary, and will have to be revisited later on. In the following chapters, I will develop a theory of clause-typing that assumes extra-compositional conventions of use, and does not make any direct use of illocutionary concepts. Insofar as it is successful, it will demonstrate that we can do without appeal to illocutionary acts. Once this theory is in place, we can reconsider the question whether there is a reason to represent force within the system of semantic composition, a point I will return to in Section 6.5.

Chapter 4

The sentential force of declaratives

In the last chapter, I raised some basic questions concerning clause typing, arguing that each clause type must be conventionally associated with its own use. In so doing, however, I only spoke in the most general terms about what such a conventionally-specified use is. In this chapter, I want to remedy this, focussing on the case of declaratives.

In our idealized system for pragmatic reasoning, \mathcal{P}_{Sen} , declaratives correspond to the sentences of the sublanguage \mathcal{L}_+ , defined as:

$$(4.1) \quad \mathcal{L}_+ := \{ \vdash \phi \mid \phi \in Prop \}$$

Since we are assuming that matrix sentences have the same denotation as embedded sentences have, we take each sentence of Sen to denote the set of possible worlds that the corresponding $Prop$ sentence denotes:

$$(4.2) \quad \llbracket \vdash \phi \rrbracket^{Sen} := \llbracket \phi \rrbracket^{Prop}$$

The question is now: What needs to be added to \mathcal{P}_{Sen} to enable the system to model the use of declarative sentences? This chapter explores—in an informal manner—a number of hypotheses that answer this question. I will settle on a particular proposal, which then is integrated into the system of dynamic pragmatics in Chapter 5.

Before I examine some of the possibilities, it will be useful to have a set of explanatory targets—things about declaratives that we want to explain. I summarize these in the rest of this section.

Declaratives are well-suited to express beliefs, and thereby to inform. This property is instrumental in explaining why declaratives are typically used to convey information: When a speaker utters a declarative, by and large, the audience is licensed to infer that he takes the content of the declarative to be true, and if the audience further takes the speaker to be well-informed, they can further conclude that the content of the declarative is indeed true. In Chapter 2, I showed how to derive such an inference in the obvious way: If the addressee believes that the speaker would utter a declarative with content p only if he believes p to be true, then observing an utterance of a declarative with content p will lead him to conclude that the speaker does, indeed, believe p to be true. But in Section 3.1, I pointed out that the contextual assumption driving this inference needs justification. And the justification must be particular to declaratives, as sentences of other types are not typically used to express beliefs.

Declaratives make claims, and thereby are subject to reprobation. When a speaker utters a declarative, he opens himself up to the possibility of reprobation. A speaker can be criticized, or be said to be ‘in the wrong’ if it transpires that he uttered a declarative he knew to be false.

One may well wonder if this is something a linguistic theory of language use should explain. Isn’t an injunction against lying rather something that should be explained by a theory of ethics or the like? It is, and that means that this property should not be explained by a linguistic theory of language use *alone*, but rather only in tandem with a suitable theory of moral behavior. But the property also cannot be explained by a theory of moral behavior alone. The theory of language use must provide a ‘hook’ on which an injunction against lying can be hung. The comparison with imperatives is useful. Whatever their precise semantics is, intuitively, the content of an imperative should determine at least *fulfillment*

conditions, and thereby determine a set of truth conditions: The conditions that obtain if the imperative is fulfilled. But when a speaker utters an imperative whose truth-conditions are not fulfilled, he does not open himself up for reprobation, and the injunction against lying does not apply. By contrast, this injunction does apply to the truth conditions associated with declaratives. But it seems odd to assume that a theory of moral behavior makes reference to morpho-syntactic distinctions. The division of labor should be roughly as follows:

- (4.3) Theorem of a complete theory of language and language use:
A speaker who utters a declarative with content p under conditions C makes a claim that p .
- (4.4) Theorem of a complete theory of moral behavior:
It is morally objectionable to make a claim that p if one does not know/believe/have good reason to believe that p (except if conditions C' obtain).¹
- (4.5) From (4.3) and (4.4), it follows that:
It is morally objectionable to utter a declarative with content p under conditions C if one does not know/believe/have good reason to believe that p (except if conditions C' obtain).

(4.5) should no more be a theorem of a theory of moral behavior alone than it should be a theorem of a theory of language and language use alone.

It is important to note that this property is, in principle, independent from the fact that declaratives are well-suited to express beliefs. Not every explanation of why declaratives have the latter property also explains why they have the former property. In many typical cases, the two properties may be argued to be tightly connected. Suppose an utterance of a declarative is taken to be honest (i.e., based on a belief that its content is true), but actually is not. In this case, we could explain

¹The set of conditions C' is a qualification necessary because I do not want to claim that the correct theory of moral behavior categorically proscribes lying in all situations.

the possibility for moral judgement in terms of norms on cooperative behavior—i.e., by an injunction to not make other people believe something that you take to be false. But a speaker is subject to reprobation for (knowingly) false statements even in contexts in which he is not presumed to be honest—if we are in a situation in which I take it to be likely that you would say things that are not true, that does not give you license to utter untruths with impunity. A speaker cannot (at least usually) argue that his declarative utterance was not a lie on the grounds that he knew in advance that the audience took him to be untrustworthy.

Utterances of inconsistent declaratives require retraction. Speakers are generally expected to maintain consistency between the contents of the declaratives they utter. This does not mean that they cannot acknowledge that they now take the content of a previous declarative to be false, and assert the opposite. But if a speaker wants to (sincerely) utter a declarative that is inconsistent with his previous ones, he has to acknowledge this fact, and generally will be required to indicate which previous declarative he wishes to retract if there are multiple such candidates.

Appreciating this property is somewhat complicated by the fact that speakers are not always *aware* of the inconsistency of two propositions. A typical instance is a failure to be logically omniscient. Suppose a group of novice students of set theory do not know that the Axiom of Choice and Zorn's lemma are equivalent, given the other axioms of ZF. About a given extension E of ZF, one of them first utters (4.6), and a short while later, he utters (4.7).

(4.6) E entails the Axiom of Choice.

(4.7) E does not entail Zorn's lemma.

His second utterance may appear to be perfectly fine, without any acknowledgement of contradiction. Pre-theoretically, we do not know whether that is because the consistency requirement does not apply in this case (because the consistency requirement really is 'Utterances that are *known* to be inconsistent require retraction') or because, even though the consistency requirement applies, the interlocutors are

unaware of what it mandates. But it is clear that if the interlocutors are or become mutually aware that (4.6) and (4.7) are inconsistent, one of the two utterances must be retracted.

Similar issues are raised by the fact that speakers and listeners do not generally have perfect recall: They do not always remember all declarative utterances that were made. So if (4.6) was made hours prior to (4.7), the inconsistency may not be recognized simply in virtue of the fact that none of the discourse participants remembers that (4.6) was uttered. In what follows, I will ignore issues of imperfect recall and failures of logical omniscience. Indeed the idealized representation of beliefs in the system set up in Chapter 2 takes agents to be logically omniscient and to have perfect recall. I will maintain this representation throughout the dissertation, as the drawbacks associated with these idealization are not at issue.

Again, I want to note that while this property often is closely connected to the fact that declaratives are good vehicles for expressing beliefs, this does not mean that any explanation of the latter property explains the need for retraction. One reason is that beliefs change over time. It is perfectly reasonable to believe, at time t , that p and to no longer believe p at a later time t' . But then, expressing (at t) the belief that p and expressing (at t') the belief that $\neg p$ is a perfectly sensible thing to do, as well. So the fact that declaratives tend to express the belief that their contents are true is not, in itself, an explanation why subsequent utterances of declaratives need to be consistent, or be made consistent by retraction.

Summary The properties we wish to explain are summarized in (4.8).

- (4.8)
- a. Declaratives are well-suited to express beliefs, and thereby to inform.
 - b. Declarative utterance make claims, which are subject to reprobation.
 - c. Utterances of multiple declaratives with inconsistent contents require retraction.

4.1 Mapping the territory

It will be helpful to have a rough understanding of the territory we are about to enter—what is the space of plausible hypotheses about sentential force? That space may well be infinite, but I will restrict attention to a number of prominent, viable hypotheses from the literature, which have been proposed either directly as hypotheses about *declaratives* or as hypotheses about the high-level speech act of *assertion*.

The accounts I will consider fall into two broad classes: On the one hand, there are accounts that take the specification of force to be a *descriptive fact* about speakers of the language, a fact about how speakers actually *do* behave. On the other hand, there are accounts that take the specification of force to be a *normative fact*, a fact about how speakers *ought to* behave.²

The first kind of account generally links declaratives to belief directly: Either it is simply taken to be a descriptive generalization that speakers tend to utter only those declaratives that they believe to be true (as in Lewis's (1975) influential account, discussed in Section 4.2) or it is said that declaratives 'express beliefs'. To be interesting, such a claim must be combined with an account of what it is to express a belief, and much of the discussion in Section 4.3 will center around this question. Ultimately, I will argue that such expressive theories are equivalent to, or need to be combined with, either a Lewis-style account or a normative account in order to explain how sentential force is determined.

The second kind of account—which takes sentential force to be essentially normative—comes in two sub-types: Those that take sentential force to be determined by normative *preconditions* (discussed in Section 4.5.1) and those that take it to be determined by normative *consequences* of utterances (discussed in Section 4.5.2). The former accounts specify 'rules' for when a declarative utterance can be made (in analogy to a law that says that one can become president of the USA

²This division roughly corresponds to Harnish's (2005) distinction between Gricean and Austinian theories—though the former include only such theories that make essential reference to the speaker's *intentions*, which arguably is not a necessary feature of the first type of theory I discuss. One prominent such theory, that of Bach and Harnish (1979), discussed in Section 4.3.1, however, does put intentions center stage.

only if one is a natural-born citizen), while the latter type of account specifies ‘obligations’ that arise from making a declarative utterance (in analogy to the action of signing a rental contract, creating the obligation to pay a certain amount of money in regular intervals).

Existing normative accounts, especially of the second subtype, relate to the notion of belief only in a rather indirect way—the normative consequences of declarative utterances do not in themselves involve the notion of belief. They only make it so that an agent who wants to abide by these obligations generally has the incentive to utter declaratives whose contents he believes to be true.

The account that I will ultimately propose (explicated in Section 4.6) can be viewed as a hybrid of ‘expressive’ and ‘normative consequences’ accounts. It directly appeals to the notion of *belief* in that it takes the conventional force of declaratives to be to create a *public belief* of the speaker, where what a public belief is is explicated in terms of its normative consequences.

This account will, in many ways, be similar to the popular account of assertion of Stalnaker (1978, et seq.). Stalnaker takes an assertion of a proposition p to be a proposal to add p to the CONVERSATIONAL COMMON GROUND of the discourse participants. This conversational common ground is understood as the set of propositions that the speakers mutually take each other to believe, or pretend to do so. In this sense, it is a notion of ‘public belief’. In the recent linguistics literature, various authors have followed Gunlogson (2003) in constructing the common ground of a set of discourse participants from their individual ‘public beliefs’ (Gunlogson 2008, Davis 2009, Farkas and Bruce 2010, Farkas and Roelofsen forthcoming, Davis 2011)—a proposition is in the common ground if it is a discourse commitment of all interlocutors. This separation is necessary to talk about cases of overt disagreement: If A asserts p and his addressee B fails to accept p as true, p is a public belief of A , but not of B . This shows that we need a notion of *individual* public belief.

It is worth noting that speaker presupposition, at least in the way Stalnaker understands it, is *not* this notion. While speaker presupposition (in contrast with the

common ground) is an individual notion, it is not suitable to model overt disagreement: “one presupposes that ϕ only if one presupposes that others presuppose it as well” (Stalnaker 2002, p. 701). It is not quite clear what ‘individual public belief’ would amount to in Stalnaker’s terminology. The only viable candidate seems to be that ϕ is an individual public belief of A iff it is common ground that A believes p . But the relationship between (actual) belief and speaker presupposition/common ground in Stalnaker’s model is not quite clear, so it is not clear what this would amount to. Also, note that this definition entails that if C and D explicitly agree that A , who is absent, believes p , then p would thereby be a ‘public belief’ of A —but that seems wrong.

My account of the sentential force of declaratives will be founded on a notion of individual public belief, which can be construed as a generalization of the notion of a ‘discourse commitment’ of Gunlogson, Davis, Farkas, et al. Besides adopting the separation of common belief into individual public beliefs, it departs from Stalnaker’s conception in that it stresses the *normative character* of ‘public beliefs’.

I shall not argue here one way or another whether the conception I will propose is an elaboration of Stalnaker’s, a variant of it, or a replacement for it. The conceptions are similar in many ways, but, minimally, they differ in emphasis, and the things I emphasize will be instrumental in what I want to explain.

4.2 Declaratives as governed by Lewis-conventions

Our aim is to add something to the denotation of declaratives, in order to explain why declaratives are well-suited for informing, while interrogatives, etc. are (usually) not. Perhaps we do not need to add much. Perhaps it is just *precedence* that enables addressees to make the right inferences. Speakers, by and large, use declaratives when they want to inform, and they use interrogatives, by and large, when they want to seek information, and so on. Addressees are aware of this, so they will assume, at least by default, that a speaker who utters a declarative (interrogative) intends to convey that he takes the declarative to be true (intends to inquire which

of the answers to the interrogative are true).³ If we adopted this view, we could treat declaratives in the way we treated the formulas of *Prop* in Chapter 2: We would simply assume that agents generally believe each other to only utter declaratives that they believe to be true (or that they desire the addressee to believe to be true), and we could add similar contextual beliefs about interrogatives, imperatives and so forth. These beliefs would be justified by the knowledge that speakers have adhered to the relevant generalization in the past.

This is essentially the view of Lewis (1975, elaborating on Lewis (1969)). He starts from the assumption that a (declarative) *language L* is simply a function from linguistic forms to sets of possible worlds (or any other specification of truth-conditions). Given this conception of a language, he asks what has to be the case for a particular language to be in use in a population.

Lewis' answer is that there has to be regularity in behavior in the population in question, such that members of the population, in general, only utter sentences of *L* if their *L*-truth-conditions obtain. At the same time, there is a regularity of behavior in that population according to which its members, generally, when they observe an utterance, assume that the *L*-truth-conditions of the sentence obtain.

Lewis requires that the regularity in question has a number of particular properties, the most important is that it be known to the members of the community, that it be *beneficial* (most members prefer that it be in place to it not being in place, most of the time), and that it be *stable* and *self-sustaining*: Given that most members of the population obey the regularity, most of the time, most members of the population have incentive to continue obeying the regularity, most of the time. Regularities in behavior that have this property (and a number of others) are what Lewis calls *conventions*. So Lewis says that a population *uses* a particular language *L* if there is a CONVENTION OF TRUTHFULNESS AND TRUST with respect to *L* in the population:

“My proposal is that the convention whereby a population *P* uses a

³Arguably, such precedential facts are also at play in the bread example in Section 3.2: While it is true that German bread is well-suited for providing delicious sustenance, there are, of course, many other uses it could in principle be put to. Similarly, the fact that it is not common practice to store one's possessions in someone else's room is instrumental in my inference that the bread is intended as a gift, and so forth.

language L is a convention of *truthfulness* and *trust* in L . To be truthful in L is to act in a certain way: to try and never utter any sentences of L that are not true in L . To be trusting in L is to form beliefs in a certain way: to impute truthfulness in L to others, and thus to tend to respond to another's utterance of any sentence of L by coming to believe that the uttered sentence is true in L ."

(Lewis 1975, p. 7)

Let us consider briefly how this conception accounts for the three explanatory targets set above, which I repeat in (4.9).

- (4.9) =(4.8)
- a. Declaratives are well-suited to express beliefs, and thereby to inform.
 - b. Declarative utterance make claims, which are subject to reprobation.
 - c. Utterances of multiple declaratives with inconsistent contents require retraction.

Lewis' account does well on (4.9a): If a Lewis convention of truthfulness is in effect, it is clear why listeners (at least by default) would assume that speakers believe the contents of their declaratives.

But it is not quite clear how well a Lewisian account does on (4.9b). Certainly, the Lewisian could appeal to communicative reasons for reprobation, appealing to a norm of cooperative/social/moral behavior like (4.10).

- (4.10) One should not: Make anyone believe p if p is not true / one does not believe that p .

But this will not be enough: It does not explain, for example, why a speaker who is criticized for uttering an untruth cannot defend himself by saying that he knew in advance that he would not be believed. That is, 'making someone believe that p ' is not general enough as a 'hook' to hang an injunction against lying on. If the

Lewisian is to explain the possibility for reprobation in general, he would have to stipulate something like (4.11).

- (4.11) One should not: Make a declarative utterance that, in general, would only be made if one believed that p if one does not believe that p .

It is not quite clear how (4.11) could be motivated—except perhaps by observing that any time someone violates the Lewis-convention of truthfulness, he thereby threatens the very existence of the convention. Given that the convention is instrumental in ensuring successful communication in many cases, this would be dangerous behavior. But this explanation makes direct reference to the convention itself, and to its communicative function.⁴ But then, it seems that the same kinds of considerations should apply to all Lewis conventions, and their functions, and so we should hypothesize a general normative rule or convention that says:

- (4.12) One should not: Violate a Lewis convention (a regularity in behavior in a population that is stable, beneficial, etc.).

But (4.12) pretty much turns Lewis conventions themselves into normative conventions. Lewis' concept of convention-as-regularity would still be useful in explaining how such conventions can arise without a pre-existing language (which was one of his major goals), but it is not clear that the concept plays any essential role once the Lewis convention has been fully established and hence 'normativized'.

Finally, with respect to observation (4.9c), it seems that the Lewisian has to rely on communicative reasons again: If a speaker uttered p at a previous time, and was believed to speak truly, but now has changed his mind and wants to assert $\neg p$, his communicative goals may well be thwarted if he does not acknowledge the falsity of his prior utterance. If he does not, his audience might assume that he believes both p and $\neg p$ (i.e., that he has inconsistent beliefs), or that he forms beliefs rather nilly-willy, and hence that his new belief in $\neg p$ is not particularly likely to

⁴And it seems that (4.11) should, as well: It does not seem right to assume that (4.11) applies to *any* action that, in general, is only be performed if the speaker has a particular belief—we do not usually fault agents for being exceptions to reliable generalizations that support belief ascriptions.

accord with the facts. Acknowledging the falsity of his previous utterances may serve to guard against audience inferences like these.⁵ So the Lewisian can explain the need for retraction when it is motivated on communicative grounds—but only then.

Summary An account of the force of declaratives in terms of Lewis-conventions explains why declaratives are well-suited to express beliefs and to inform. It also explains why a speaker is subject to reprobation for utterances of declaratives that he knew to be false—but only if he had reason to think that his utterances will be taken to be true. Similarly, it can explain why retraction of inconsistent declaratives is necessary—but only if this can be motivated by communicative reasons. At first glance, this may seem sufficient, but as I will argue in Section 4.7.4, it is not.

4.3 Declaratives as expressing beliefs

Starting from our first target for explanation—declaratives are well-suited for expressing beliefs—an obvious idea is to assume that utterances of declaratives simply *are* expressions of belief. Such an analysis would assume a rule, convention, or principle that ensures that (4.13) is true.

(4.13) Whenever a speaker utters a declarative with content p , he thereby expresses a belief that p .

As will become clear in the following subsections, I think that such an approach is ultimately unhelpful in understanding declarative (or any other) sentential force. I shall discuss it at some length because it is a popular and initially attractive way to answer the question ‘What do declaratives do?’—they express beliefs! We shall see, however, that this thesis, in the way it has been understood, either (a) is untenable, or (b) is essentially Lewisian, or (c) is essentially a variant of the normative theories to be discussed in Section 4.5.

⁵Observe, though, that in many cases, *only* an acknowledgement of falsity is needed—not an exhaustive explanation why the previous belief was justified, and why the new belief is justified.

(4.13) immediately raises the question what ‘expressing’ a belief amounts to. I will discuss three possible answers in the rest of this section (and a fourth in Section 4.4). The first two can be set aside quickly; the third requires more substantive discussion.

Hypothesis I: To express a belief is to indicate that one has it. This conception is essentially the broad sense of ‘expressing’ I have used when I said that declaratives are well-suited to express beliefs, meaning that they are well-suited to give one’s audience a reason to think that one has the belief.

On this conception, saying that declaratives, by convention, express beliefs is simply saying that utterances of declaratives are actions that, by convention, are generally good reasons to think that the speaker believes the sentence to be true. The obvious question is why that would be so, and the obvious answer is: Because that is what speakers tend to do. On this conception, then, the thesis that declaratives express the belief that their content is true is simply Lewis’ idea of a convention of truthfulness, and everything I have said above applies to the expressive thesis, as well.

Hypothesis II: To express a belief is to perform an action caused by it. Williams (2002) and Owens (2006) analyze assertions as expressions of belief, with a rather different conception of what ‘expressing’ means. According to their conception, to say that an action is expressing a belief is to say that it is caused by a belief. Among other things, that means that an action can be said to express a belief only if the agent in question actually *has* the belief. Thus Owens writes:

[E]xpressing a belief in action is one way of *acting on* that belief and one can’t act on a belief which one does not have. So, as I use the term, an expression of belief may be poor or inadequate but it can’t be insincere: you can’t express beliefs you don’t have.

(Owens 2006, p. 109)

This sense of ‘expressing’ is rather liberal: Any action that is motivated by a certain belief counts as expressing that belief. As MacFarlane (2011, p. 81) puts it: “In reaching for the umbrella as I head to the door, I express my belief that rain is likely.”⁶ But it is not liberal enough. As an account of assertion, it is too restrictive because it rules out the possibility of insincere assertions.⁷ And as a claim about declaratives (‘Whenever a speaker utters a declarative, he performs an action caused by a belief in the truth of its content’), it is simply false. The only way to save such an account of declaratives would be to say, instead, that an utterance of a declarative is only *sincere* if the utterance is caused by the appropriate belief. But this just begs the question (see also Section 3.5). Without either a descriptive generalization that speakers tend to speak sincerely (i.e., a Lewis-convention of sincerity) or a normative rule that speakers ought to speak sincerely, this version of the expressive thesis does not have any explanatory power.

I conclude that the Williams (2002)/Owens (2006) conception of ‘expressing a belief’ is not useful in elucidating the conventions governing the use of declaratives, for it is either hopeless or it ends up being equivalent to one of the alternative conceptions explored in this chapter.

Hypothesis III: To express a belief is to have a reflexive intention. The rest of this section will be concerned with a quite elaborate, and somewhat popular, conception of what it is to express a belief (or an attitude in general), viz., that of Bach and Harnish (1979). Their account is somewhat exceptional in philosophical speech act theory in that they explicitly relate their account of assertion in terms of belief expression to sentence types, in particular with declaratives (though, as we will

⁶I am not certain that this use of **express** is consistent with its use in everyday conversation—such a taking of an umbrella certainly can be described, on occasion, as expressing the belief that it is raining—especially if it is done deliberately with the intention to convey this belief—but it is not clear to me that any taking of an umbrella that is motivated by a belief that rain is likely counts as ‘expressing’ that belief.

But we can, of course, take **express** to be a technical term here, or we could replace it by another, such as **display**, without changing anything about the putative analyses of declaratives and assertions.

⁷This leads Williams (2002, p. 74) to define assertion disjunctively: “A asserts that *p* where A utters a sentence *S* which means that *p*, in doing which either he expresses his belief that *p*, or he intends the person addressed to take it that he believes that *p*.”

see below, this connections turns out to be an essentially Lewisian one). This is the first reason I discuss their account at some length. The second is that their intricate account poses some conceptual difficulties which make it hard to evaluate. Since they are also the main proponents of a popular approach to explicitly performative sentences, to be discussed in Chapter 7, it is useful to point out these conceptual difficulties here.

4.3.1 Bach and Harnish's (1979) on communicative speech acts

Unlike Austin and Searle, Bach and Harnish (1979) (B&H) sharply distinguish two kinds of illocutionary acts, following Strawson (1964): Those that are essentially conventional and those that are, by contrast 'communicative'. The former category contains EFFECTIVES such as bidding and bequeathing, as well as VERDICTIVES, such as finding guilty. The latter class contains most other illocutionary acts, in particular assertions, promises, orders, etc. According to B&H, such communicative speech acts uniformly 'express' attitudes, such as beliefs, desires, and intentions. Their conception of 'expressing' is based on the notion of a reflexive intention, or R-intention, which is familiar from Grice's (1957) definition of 'non-natural meaning':

(4.14) To R-intend is "to produce some effect in an audience by means of the recognition of this intention" (Grice 1957)

In terms of this, B&H defines what it is to express an attitude:

(4.15) EXPRESSING: For *S* to *express* an attitude is for *S* to R-intend the hearer to take *S*'s utterance as reason to think *S* has that attitude.
(Bach and Harnish 1979, p. 15)

To express a belief thus amounts to (4.16).

(4.16) EXPRESSING A BELIEF THAT *p*: For *S* to *express* the belief that *p* is for *S* to R-intend the hearer to take *S*'s utterance as a reason to think that *S* believes that *p*.

Unfortunately it is unclear how (4.16) should be understood. What exactly does it mean to 'R-intend the hearer to take *S*'s utterance as a reason to think that *S* believes that *p*'? We need to understand this in order to be able to evaluate the thesis that declaratives express beliefs against our three explanatory targets:

- (4.17) =(4.8)
- a. Declaratives are well-suited to express beliefs, and thereby to inform.
 - b. Declarative utterance make claims, which are subject to reprobation.
 - c. Utterances of multiple declaratives with inconsistent contents require retraction.

The most important question in this connection turns out to be: Is to R-intend that *Y* also to intend that *Y*? That is, does someone who expresses a belief that *p* intend his addressee to think that he has the belief that *p*? As we shall see, the answer to this question is not clear.

4.3.2 R-intentions à la Bach & Harnish

For Grice, to mean ('declaratively') that *p*, roughly, is to intend to get *H* to form the belief that *p*, by means of recognition of the intention itself. This makes quite plain that to Grice-R-intend that *Y* is (in part) to intend that *Y*.

Things are not so clear for B&H, though. They find fault with Grice's definition insisting (following Searle (1969, p. 46–50)) that 'to get *H* to believe that *p*' (or their equivalent 'for *H* to take *S*'s utterance as a reason to think that *S* believes that *p*') is a mere *perlocutionary* effect that does not need to be satisfied for *S*'s illocutionary intention to be fulfilled.

"Searle argues that the sorts of effects Grice mentions, such as beliefs, intentions, and actions, are not produced by means of recognition of the intention to produce them. For example, the hearer might recognize that he is to believe something and yet refuse. These sorts of effects are

perlocutionary, and the speaker's illocutionary act, whose identity he is trying to communicate, can succeed without the intended perlocutionary effect (if there is one) being produced. So a reflexive intention is involved in communication, just as Grice claimed, but the kinds of intended effects he specified are not of the right sort. Getting the hearer to recognize them does not constitute producing them. In section 1.6 we consider just what sort of reflexive intention is fulfilled merely by being recognized."

(Bach and Harnish 1979, p. 14)

Unfortunately, B&H never define what it is to 'R-intend that *Y*', but they make various remarks about the notion:

[T]he intended effect of an act of communication is not just any effect produced by means of recognition of the intention to produce a certain effect, *it is the recognition of that effect.*

(Bach and Harnish 1979, p. 15, emphasis in original)

And, after introducing their definition of 'expressing an attitude' ((4.15) above), they write:

Accordingly, the intended illocutionary effect (or simply illocutionary intent) is for H to recognize that R-intention.

(Bach and Harnish 1979, p. 16)

So B&H stress that the intended effect is the recognition of the intention itself, and that the R-intention is "fulfilled merely by being recognized". But if the intention is fulfilled by being recognized, it seems it *cannot* involve an intention for something *other* than being recognized. But then, it seems to 'R-intend that *Y*' cannot involve an intention to *Y*—for such an intention is fulfilled only if *Y* obtains, not just if the intention is recognized. But then, it becomes difficult to see what an 'R-intention

to Y' really could be—what is the role of Y in the intention, if it is not the thing that needs to occur for the R-intention to be fulfilled?

An intention that is fulfilled once it is recognized, is, it seems just an intention to be recognized—adapting Searle (1969)'s helpful notation, in which every occurrence of the verb **intend** or the noun **intention** is indexed with a name for the intention predicated, this amounts to the statement in (4.18):

(4.18) S R-intends iff S intends $_{i1}$ that his intention $_{i1}$ is recognized.

But what this formulation makes obvious is that there is no place for an argument of 'R-intend' in the *definiens*: According to this definition, there can only be *one* R-intention, whose content is only its own recognition. We can contrast this with Grice's original formulation (somewhat expanded), in the same format:⁸

(4.19) S R-intends Y iff S intends $_{i1}$ that

- a. Y ;
- b. the intention $_{i1}$ is recognized;
- c. Y obtain, in part, because of (b).

But (4.19) will be unsuitable for B&H, it seems, because it involves an intention to Y , and hence the intention described in (4.19) will not be fulfilled merely by being recognized (this only satisfies part (b)), but requires further that Y obtain.

At the same time (4.18) will not be useful in communication, because (4.18)

⁸Considering this version of Grice's definition helps to foreground that the conceptual difficulty I am seeing here is *not* the one raised by Sperber and Wilson (1986), which casts doubt on *any* notion of a reflexive intention: Sperber and Wilson argue, in a nutshell, that reflexive intentions cannot be mentally represented, due to their self-referential nature. Their argument involves the premiss that, in order to mentally represent an intention like (4.19), the occurrence of 'the intention $_{i1}$ ' in (b) has to be replaced by a representation of $i1$ itself, leading to an infinite regress. I think this argument can be resisted, on the assumption that such a mental representation could involve 'mental demonstrative', or something of the kind (essentially, a pointer) that refers to the overall intention i_1 . Siebel (2003) mentions such a possibility, but expresses concerns about such a mental demonstrative device. It may not be trivial to posit the existence of such mental representations, but given modern computer languages are rife with self-referential representations (functions that call themselves, objects whose members are objects that point to the parent object, etc.), it does not seem too far-fetched to assume that mental representations could have the same feature.

carries only one bit of information—if understanding were the recognition of R-intentions as characterized in (4.18), then to understand someone would simply be to say ‘Ah, he R-intends!’—not ‘He R-intends me to *Y*’, or anything like that. So it seems (4.18) cannot be what B&H have in mind.⁹ But at the same time, if their R-intentions really are entirely fulfilled just by being recognized, it seems that (4.18) *must* be what they had in mind.

I think Siebel (2003) is getting at the same conceptual difficulty when he discusses what he calls a ‘mereological difficulty’ (which he attributes to Wayne Davis (p.c.)):¹⁰

“Second, [R-intentions] give rise to a mereological difficulty. Let us abbreviate the content of a particular reflexive intention by ‘*X*’: *S* expresses the attitude *A* iff *S* intends *X*. This intention is supposed to be identical with the intention that an addressee *H*, by means of recognizing that *S* intends *X*, take the utterance as a reason to think *S* has *A*. Intentions, however, are individuated by their content. That is, intending *p* is identical with intending *q* only if ‘*p*’ and ‘*q*’ specify the same content. Hence, the content given by ‘*X*’ should be identical with the content given by ‘*H*, by means of recognizing that *S* intends *X*, takes the utterance as a reason to think *S* has *A*’. But how could they be identical? After all, the former content seems to be a proper part of the latter because the sentence specifying the latter contains ‘*S* intends *X*’ as a proper and semantically relevant part.”

⁹(4.18) has *one* advantage: It makes sense of B&H’s claim that R-intentions are “intended to be recognized as intended to be recognized”. Siebel (2003, p. 357) notes this, wondering whether it is supposed to follow from the definition of an R-intention. According to (4.18), R-intentions indeed are intended to be recognized as intended to be recognized—and only that.

¹⁰The only place I know of where Siebel’s ‘difficulty’ has been addressed is Witek (2009, p. 77). But Witek proposes the following description of the content (again with Searlean indexing):

- (i) *S* intends_{*i*1} that *H*, by means of recognizing this intention_{*i*1}, takes *S*’s utterance as reason to think that *S* has *A*.

But (i) is like Grice’s definition in that it is not an intention merely to be recognized, but rather an intention that is fulfilled only if the ‘perlocutionary effect’ (*H* takes *S*’s utterances as a reason to think that *S* has *A*) occurs.

(Siebel 2003, p. 356)

But humbly calling the issue a ‘mereological difficulty’ obscures its importance: One might think that all Siebel’s worry shows is that the part-whole structure of intentions (or their contents) is quite special, in that it does not validate all principles that other mereological structures satisfy. I think, to the extent that the considerations above are correct, they actually show that B&H’s conception of R-intentions is *incoherent*. This would be quite devastating, as R-intentions are the very foundations on which their whole theory is build.

Perhaps we can salvage their notion of an R-intention, if we interpret their characterizations of what it takes to be an illocutionary R-intention somewhat loosely. Here is a version that is, I believe, coherent, and that has some of the properties B&H desire. I am not sure it is what they had in mind, but seeing as it is the only coherent interpretation that I have been able to come up with, it is the one that I am going to work with in what follows:

- (4.20) S R-intends that Y iff S intends _{i_1} that
- a. H recognizes that S intends _{i_2} that Y ;
 - b. H recognizes the overall intention _{i_1} ;
 - c. (a) obtains, in part, in virtue of (b).

Now, the intention in (4.20) does not quite fit B&H’s description. The intention i_1 is not fulfilled just by being recognized: For i_1 to be fulfilled, *another* intention has to be recognized, as well, namely i_2 , the intention to Y . This intention itself, however, is not *part* of the intention i_1 , it just occurs in its content.

If we now assume in addition that ‘recognize’ in the description of the intention is factive—and it seems that B&H assume this—then S cannot intend i_1 without also having the intention i_2 . But then, H cannot ascribe i_1 without also ascribing i_2 and so recognizing that S has i_1 will involve recognizing that he has i_2 , as well. So, once i_1 is recognized, i_2 must have been recognized as well.

On this interpretation, R-intending that Y is not intending that Y —the R-intention can be fulfilled without Y being true, but it still is the case that one

cannot R-intend that Y without also intending that Y . Again, I am not sure that this is what B&H have in mind,¹¹ but it is the version I am going to work with in the sequel.

4.3.3 Expressing a belief as R-intending

With a workable understanding of an R-intention, we can now consider the hypothesis that declaratives express beliefs, on my reconstruction of B&H's definition of an R-intention. We then can evaluate how well this hypothesis fares in helping us meet the explanatory targets set out at the beginning of this chapter. I repeat their definition of 'expressing a belief':

- (4.21) EXPRESSING A BELIEF THAT p : For S to *express* the belief that p is for S to R-intend the hearer to take S 's utterance as a reason to think that S believes that p .

With this, the principle (4.13) becomes (4.22):

- (4.22) Whenever a speaker utters a declarative with content p , he thereby R-intends the hearer to take his utterance as reason to think S believes that p .

This immediately raises the following worry: Declaratives are often uttered, even sincerely, if the speaker does not, and cannot, intend that his audience form the appropriate belief—e.g., if the speaker knows that he will be not be trusted, but wants to 'go on record' with his statement.

But it is worth setting this worry aside for a moment to dwell on B&H's conception. Recall again that B&H's claim is that *asserting* is the expression of a belief.

¹¹One reason to doubt that this is exactly what B&H have in mind is that they appear to think that intending that ' H take S 's utterance as a reason to think he has the attitude' is a content that is somehow particularly suited for R-intentions. But the definition in (4.20) does not put any requirements on Y —any Y could be R-intended according to this definition, as long as it is something that S can intend.

So maybe it is not that B&H's conception of expressing works as a way to understand the thesis that (all) declaratives express beliefs, but maybe their conception of how *asserting* works (when it does) with declaratives can provide a window into understanding declaratives—for unlike other authors mainly concerned with the illocutionary act of assertion, B&H do spell out how asserting relates to declarative force—to an extent.

They relate illocutionary acts to the content of declaratives by the conception of *literalness of a speech act*. They write $\vdash (\dots p \dots)$ to represent 'what is said by a declarative' (where $(\dots p \dots)$, roughly, characterizes truth conditions and \vdash is just a representational device), and then state the following COMPATIBILITY CONDITION, which defines the notion of L-compatibility (Bach and Harnish 1979, p. 34): An utterance is L-compatible if . . .

(4.23) If S is saying that $\vdash (\dots p \dots)$, S is expressing a belief that p .

And then they use this notion to define what a literal performance is (where F is an illocutionary act type, $*$ a metavariable over symbols such as \vdash):

(4.24) *Literal performance* (Lit) S 's F -ing that P in saying that $*(\dots p \dots)$ is literal just in case:

- i. $P = (\dots p \dots)$
- ii. F -ing is L-compatible with saying that $*(\dots p \dots)$

What this boils down to, in the case of declaratives, is essentially this:

(4.25) An utterance of a declarative with content p is LITERAL if it expresses a belief that p .

Spelling out Bach and Harnish's conception of expressing:

(4.26) An utterance of a declarative with content p is LITERAL if its speaker R-intends his audience to take his utterance as a reason to think that he believes p to be true.

But this is not essentially different from the hypothetical specification of sincerity conditions in Section 3.5. And the same considerations apply: (4.26) does not tell us anything about how declaratives are used, unless it is paired with an assumption to the effect that speakers generally do, or generally should, strive to be literal. B&H actually articulate such a principle:

(4.27) *Presumption of Literalness (PL)*: The mutual belief in the linguistic community C_L that whenever any member S utters any e in L to any other member H , if S could (under the circumstances) be speaking literally, then S is speaking literally.

This is no different from saying:

(4.28) There is a Lewis convention in C_L to the effect that speakers speak literally and that hearers take speakers to speak literally.

Bach and Harnish's conception of declarative sentential force is hence essentially Lewisian—except that the *content* of the convention is slightly different. Lewis' original proposal said that speakers tend to utter declaratives only if they are true, and hearers tend to expect them to, while (4.28) says that speakers utter declaratives only if they R-intend to convey that they believe the content to be true.

This brings us back to the worry, set aside above, about whether it makes sense to assume that any speaker who utters a declarative (without any overt indication of insincerity) can be said to R-intend his audience to think he believes the semantic content of what he says. In particular, it is easy to imagine contexts in which a speaker knows that he will not be trusted (imagine, for example, a suspect being questioned by the police, who he knows think he is guilty—such a suspect may well insist he is innocent, even though he knows his interrogators will not believe that he believes himself to be innocent. But if he knows this, he cannot intend to change their mind.

It seems that what B&H must say about such a case is that the suspect is not speaking literally. Perhaps because this seems an odd thing to say (at least on

our everyday understanding of the word **literal**), Bach and Harnish argue that the speaker can still be taken to express a belief in such a case—revising their characterization of expressing a belief:

Instead of saying that expressing an attitude is R-intending *H* to take one's utterance as reason to believe that one has that attitude, we can say that it is R-intending *H* to take one's utterance as sufficient reason, unless there is mutually believed reason to the contrary, to believe that one has that attitude.

(Bach and Harnish 1979, p. 291)

As MacFarlane (2011) discusses, it is not clear that this revised explication of 'expressing' can be made sense of, as it appears to require us to believe that an agent can 'intend *p* unless *q*' if he knows that *q* is true. Suppose I concluded the discussion of B&H's account with the sentence in (4.29). What kind of intention would I claim to have?

(4.29) I intend to discuss this issue further, unless the present work is a dissertation in linguistics.

In summary, it seems that B&H's account does not fare any better (or worse) than Lewis' original account on the explanatory targets at issue here:

(4.30) =(4.8)

- a. Declaratives are well-suited to express beliefs, and thereby to inform.
- b. Declarative utterance make claims, which are subject to reprobation.
- c. Utterances of multiple declaratives with inconsistent contents require retraction.

As on Lewis' account, target (4.30a) is easily met, on the assumption that speakers typically don't intend to make their addressees believe things that they do not believe themselves. Targets (4.30b) and (4.30c) can, again as on Lewis' account

only be met in cases in which there are *communicative* reasons for reprobation or retraction: We can explain why a speaker can be criticized for speaking falsely in cases in which he has reason to think that he would be believed, but not otherwise. And we can explain why a speaker would feel the need to retract if he otherwise runs the risk of being thought of as holding inconsistent beliefs, but not otherwise. As indicated before, while that may seem sufficient at first glance, we will see in Section 4.7 that it is not.

4.3.4 Summary

The main result of the present section is this: while it may be adequate, at a certain level of description, to say that declaratives express beliefs, saying so does not help us in understanding how declaratives are used. The conceptions of ‘expressing’ I discussed are either equivalent to, or require an independent specification of, a Lewis-convention or a set of normative rules governing the expression of beliefs. I conclude that expressive accounts, at least from our current perspective, are not an *alternative* to Lewis-style accounts and normative accounts of clause typing, but rather are *variants* of such accounts. Consequently, I will largely ignore this category of account in what is to follow.

4.4 Counts-as rules

The notion of *constitutive rules* have played a significant role in philosophical speech act theory, especially in the work of Searle (1969). Constitutive rules are, in the words of Searle, rules that “do not merely regulate, they create or define new forms of behaviour. The rules of football or chess, for example [...] create the very possibility of playing such games.” For Searle, illocutionary acts are defined by exactly such rules.

There is no general agreement on how constitutive rules should be understood, or whether they exist at all (for various views on the matter and discussion, see Ransdell (1971), Giddens (1984) and more recently Hindriks (2009)). In recent

discussion (e.g., MacFarlane (2011, Section 2)), constitutive rules are often taken to have the same format as other, ‘regulative’ rules, i.e., they specify what one must or must not (or should and should not) do. These kind of rules will be the topic of Section 4.5.1.

Searle himself tends to talk about constitutive rules in terms what he calls ‘the counts-as locution’. And indeed, if we look to games such as soccer, it seems we have such rules:

(4.31) OFFSIDE RULE

A player counts as being in offside if he is nearer to his opponents’ goal line than both the ball and the second-last opponent (and is not in his own half of the field of play).

Searle stresses that such counts-as rules are not mere definitions:

“‘offside’, ‘homerun’, ‘touchdown’, ‘checkmate’ are not mere labels for [a] state of affairs [. . .], but they introduce further consequences by way of, e.g., penalties, points, and winning or losing.”

(Searle 1969, p. 36)

Counts-as rules are constitutive rules, then, because they are part of a larger set of rules that specify normative consequences of something counting as something else. Perhaps we could take this as a model, and simply say that ‘expressing a belief’ is something that is, in part, constituted by the rule governing declaratives:¹²

(4.32) DECLARATIVE RULE (counts-as version)

When a speaker utters a declarative with content *p*, his utterance counts as an expression of the belief that *p*.

We would not define ‘expressing a belief’, then, and would say that ‘expressing a belief’ is not something that is independently defined—expressing a belief is

¹²Of course, one can also read (4.32) as an empirical claim—but then we need to specify independently what ‘expressing a belief’ means, and (4.32) itself is of no explanatory value.

simply the thing that one does when sincerely uttering a declarative.¹³

It should be clear, though, that no explanatory work is done by a rule like (4.32). It will not tell us why speakers tend to utter declaratives only when they take their contents to be true, and it will not tell us why speakers of declaratives are subject to reprobation if they fail to believe their contents to be true, or why subsequent utterances of declaratives must be consistent, or made consistent by means of retraction. This work must be done by additional rules that put constraints on expressions of belief. The same is true of the offside rule in (4.31), which does not tell us what the consequences of being in offside are. It needs to be combined with other rules of soccer that spell out the normative consequences of certain actions performed in an offside position.

That is, on the ‘counts-as’ rule approach hypothesized above, the explanatory work has to be done by normative rules like the ones discussed in the next section.

4.5 Normative theories of clause typing

Neither Lewis-conventions, nor the counts-as, or constitutive rules investigated in the previous sections are, by themselves, normative. In order to explain apparently-normative facts about language use, they have to be paired with normative constraints that are motivated independently.

There is an obvious alternative: Assume that the form–force mapping itself is governed by properly normative conventions (or perhaps simply by ‘norms’—I shall use the two terms interchangeably).

This assumption is, or at least can be, independent from the thesis that ‘meaning is normative’ (as claimed, e.g., by Boghossian (1989)). Saying that what language

¹³This is not quite Searle’s view, though. His essential rule for assertions in Searle (1969) is the somewhat opaque:

- (i) [The act] counts as an undertaking to the effect that p represents an actual state of affairs. (Searle 1969, p. 66)

In Searle (1989), he says that ‘a statement is a commitment to the truth of the [stated] proposition’—indicating that anything that counts as such an undertaking gives rise to such a commitment.

users *do* with a sentence S that has truth-conditions ϕ is constrained by norms does not amount to saying that the relationship between S and ϕ itself is normative, or determined by normative facts (nor does it deny this).

Normative theories of the form–force mapping come in two kinds: On the one hand, there are theories that take the relevant norms to specify *normative preconditions* on when a sentence can be used. On the other hand, there are theories that specify *normative consequences* of utterances.

4.5.1 Normative preconditions on utterances

Accounts of this type assume normative conventions that specify when it is appropriate to utter a sentence of a particular type. If a speaker utters a sentence in violation of these conventions, he is in the wrong. Stenius' rule for declarative mood can be seen as an early version of such a theory:

- (4.33) Produce a sentence in the indicative mood only if its sentence-radical is true. (Stenius 1967, R 3, p. 268)

We find several variants on this in the philosophical literature on assertion. Here I shall draw mainly on Williamson (1996), who proposes to understand assertion as governed by a normative rule and MacFarlane (2011), who provides a useful survey, and critical assessment of, various families of philosophical theories of assertion. Williamson (1996) himself defends the *knowledge rule* of assertion, which is restated in (4.34) as a rule about declarative utterances:

- (4.34) KNOWLEDGE RULE
One must: Utter a declarative with content p only if one knows that p .

Williamson also discusses a TRUTH RULE which in terms of declarative sentences amounts to Stenius' rule above. MacFarlane (2011) adds the REASONABLE-TO-BELIEVE RULE, variants of which are defended by Douven (2006) and Lackey (2007). I again restate it in terms of declarative utterances:

(4.35) REASONABLE-TO-BELIEVE RULE

One must: Utter a declarative with content p only if it is reasonable to believe that p .

It seems intuitively obvious that the TRUTH and KNOWLEDGE rules (on the assumption that knowledge requires truth) are too strong. Adopting them would mean that a speaker who utters a declarative p is at fault if p is false, even if the speaker had good reason to believe it is true. Such a speaker is in error, but it does not seem right to say that he has violated a rule. To be sure, we sometimes, perhaps often, criticize speakers for asserting something that is false, but in these cases, we generally either allege that the speaker did not believe what he asserted (and hence criticize him for asserting something he did not believe), or we claim that he *should not* have believed what he asserted.

This does not necessarily mean that these rules are untenable: One could argue that if a speaker honestly but erroneously believed he was uttering a true declarative (or one that he knew to be true), then we are generally willing to forgive the rule violation because he *thought* he was in compliance with the rule. I shall not argue about the merits of this move, but for our purposes, this understanding makes the TRUTH rule equivalent to the REASONABLE-TO-BELIEVE rule, and the KNOWLEDGE rule equivalent to the rule in (4.36):

(4.36) One must: Utter a declarative with content p only if it is reasonable to believe that one knows that p .

But then, I think we can safely focus on the REASONABLE-TO-BELIEVE rule (as any additional constraints on knowledge will not play a role in what is to follow). How does the rule do with respect to our explanatory targets, repeated in (4.37)?

- (4.37)
- a. Declaratives are well-suited to express beliefs, and thereby to inform.
 - b. Declarative utterance make claims, which are subject to reprobation.
 - c. Utterances of multiple declaratives with inconsistent contents require retraction.

Presumably, we can assume that a speaker who strives to obey the REASONABLE-TO-BELIEVE rule will mostly only utter sentences that he believes to be true, explaining why declaratives are well-suited to express beliefs. And quite obviously, (4.37b) is directly accounted for by the REASONABLE-TO-BELIEVE rule. But it is not clear that (4.37c) is explained under this rule.

What is reasonable to believe surely can change over time, even within the space of a single conversation—why, then, do incompatible declaratives require retractions, if they are governed by a rule that says that the speaker should only utter declaratives that he has good (enough) reason to believe? Should a declarative utterance that contradicts a previous one not simply be understood as indicating that a speaker has changed his mind? Why does he need to acknowledge this inconsistency?

The TRUTH and KNOWLEDGE rules might be seen to fare better on this count, implausible as they may be on other grounds: If the speaker changes his mind about the truth of a previously-uttered declarative, this means he realizes that his previous utterance constituted a rule violation. Perhaps it is simply necessary to acknowledge this fact. This would be essentially equivalent to assuming an independent normative rule for retraction, such as the following two hypothesized by MacFarlane (2011):

(4.38) RETRACTION RULE 3:

One must: retract a previous assertion *A* when one knows that one performed *A* and that the content of *A* was untrue.

(4.39) RETRACTION RULE 4:

One must: retract a previous assertion *A* when one performed *A* and *A* was untrue.

But these rules prove too much: It is not in general the case that one must retract all utterances of declaratives (or all assertions) that one realizes were in error. Suppose that you want talk to John, who takes part in a all-day conference. Misremembering the conference schedule, I sincerely utter (4.40).

(4.40) The last coffee break starts at 5pm.

In actual fact, the last coffee break starts at 4:30 and ends at 5. But my mistake is harmless, because you have another appointment at that time anyway, and decide to bother John during the lunch break at 12. If I realize my mistake the day after, there may be no need to retract my assertion in (4.40). If you realize the mistake, you might question me about why I informed you wrongly, and if my mistake resulted in you or John being inconvenienced, I might apologize, but it is not the case that I violate a rule if I don't. In general, it seems that I need to retract only those utterances of declaratives that are such that their truth or falsity are still *relevant*—or, if they are no longer relevant, but I still have reason to talk about the actual facts.

We could, of course, adapt the retraction rules to something like (4.41).

(4.41) RETRACTION RULE 4:

One must: retract a previous assertion *A* when one performed *A* and *A* was untrue and *A* is still relevant; or if one has reason to assert something that is incompatible with *A*.

But with this, it seems that the retraction rule has become an important part of the specification of the effect of assertions/utterances of declaratives. The retraction rule states a *normative effect* of uttering declaratives, and in doing so, it supplies something that intuitively was missing in the statement of the REASONABLE-TO-BELIEVE rule. In general, it seems that rules about normative preconditions for utterances need to be complemented by rules that govern their normative effects, i.e., rules that specify what the speaker's behavior should be like *after* the utterance, not only what should be the case *prior to* or *at* the time of the utterance.

4.5.2 Normative effects of utterances

Given that we need to specify normative effects of utterances anyway, we may consider the possibility that clause-typing conventions are about normative effects

in the first place. Maybe we then can dispense with the assumption of rules specifying normative preconditions at all, or we can assume that declarative utterances are subject to *both* rules about their preconditions and rules about their normative consequences.

MacFarlane (2011) summarizes the essential features of an account (of assertion) in terms of normative consequences:

“[T]his approach defines assertion in terms of its “essential effect.” But it regards this essential effect as the alteration of a normative status—the acquisition of new commitments or obligations.

It is important to see how the commitment approach differs from “constitutive rules” approach we considered above. Both describe assertion in essentially normative terms. But, while the constitutive rules approach looks at “upstream” norms—norms for making assertions—the commitment approach looks at “downstream” norms—the normative effects of making assertions.”

(MacFarlane 2011, P. 91)

What kinds of commitments could be the result of utterances of declaratives? In authors like Searle (1969, p. 29) we find locutions like ‘commitment to the truth of *p*’, indeed “a (very special kind of) commitment to the truth of a proposition”. But this is not helpful. If our account of the normative effects of declarative utterances is to have any bite, we must specify what ‘commitment to the truth’ amounts to. MacFarlane (2005) usefully summarizes various proposals that have been made in the context of analyses of Assertions:

- (W) Commitment to withdraw the assertion if and when it is shown to have been untrue.
- (J) Commitment to justify the assertion (provide grounds for its truth) if and when it is appropriately challenged.
- (R) Commitment to be held responsible if someone else acts on or reasons from what is asserted, and it proves to have been untrue.

(MacFarlane 2005, p. 318)

I shall discuss these in turn, before offering my own conception of the normative effects of utterances of declaratives, which I have developed jointly with Cleo Condoravdi (Condoravdi and Lauer 2011, 2012).

Commitment to withdraw

(W) is similar to the retraction rule which I argued to be too strong above, but it is weaker: It only requires retraction if the assertion is *shown* to have been untrue. This may well be weak enough to not require retraction of irrelevant falsehoods, but it is also far too weak to account for anything *but* the need to retract in response to overt demonstrations that one's utterance was false.

Perhaps, if we take 'shown to have been untrue' as weak enough, if we allow that the obligation to retract does not rely on conclusive demonstration, but only on reasonable doubt, we might be able to explain why speakers tend to utter declaratives they take to be true, and hence why declaratives are good for conveying information, if we assume that speakers have a preference against retracting their own utterances.

But in any case, it seems that we want (W) to be a *result* of our account of the uses of declaratives, rather than an independent stipulation. There is no direct connection between (W) and the notion of belief, and it only partially explains the need for retraction (since it does not explain why a speaker who decides, for himself, that he now would like to assert the opposite from what he asserted earlier

must retract). (W) *could* be an independent component of the commitments we undertake when uttering declaratives, but it would be preferable if we can derive it from more basic assumptions about declarative utterances.

Commitment to defend

(J) is repeated below:

- (J) Commitment to justify the assertion (provide grounds for its truth) if and when it is appropriately challenged.

This view has been held (for assertions), for example, by Brandom (1983). It has the potential to explain why declaratives are well-suited to expressing beliefs, even in cases in which there is conflict of interest—any assertion of something that one does not believe would make one liable to mount a futile defense. And presuming that incompatible claims cannot be jointly defended, such a view might be seen as predicting the need for retraction in case of assertions with incompatible contents.

But the idea that utterances of declaratives commit their speakers to defend the content when challenged seems to presuppose too special a kind of conversational context to be generally applicable. MacFarlane (2005) articulates this worry well:

“[T]his may be over-generalizing from seminar-room assertions to assertions in general. Suppose someone were to say: ‘You’ve given some very good reasons to doubt the truth of what I asserted. I have nothing to say in answer to your objections, yet I continue to stand by my claim.’ She would not be playing the game of assertion the way philosophers play it, but perhaps philosophers do not get to set the rules here. We would surely take her assertions less seriously than we would if she were responsive to reasons. But would we cease treating her as an asserter at all? That is not so clear.”

(MacFarlane 2005, p. 318)

It is true that in many situations—familiar not only to philosophers but to any academic—we expect that speakers are willing and able to defend their assertions, and we think less of them if they cannot or won't. But it does seem too strong to say whenever someone fails to defend a claim that is challenged (even if he gives responses much less polite than the one imagined by MacFarlane—e.g., **Get lost, I won't argue with you** or **I won't stand being questioned by you** or **I believe it, regardless of what you say**, etc.) has thereby violated a commitment.

Note also that while this conception allows us to understand why speakers would *tend* to retract assertions if they want to assert something that is incompatible with them, the assertion of two inconsistent propositions would in itself not be a violation of the commitment—but it seems that if someone makes two incompatible assertions—at least when he does so knowingly—he has thereby, with his second assertion, already violated the commitment he took on with the first. But this does not follow from (J).

So (J), again, appears to only partially capture the commitments we take on with utterances of declaratives, and it also appears to predict a commitment that is often not intuitively present.

4.5.3 Commitment to be held responsible for consequences

(R) Commitment to be held responsible if someone else acts on or reasons from what is asserted, and it proves to have been untrue.

Such a view is evident, in particularly strong form, in von Savigny (1988)'s characterization of constatives:

“A conventional make-up which is fundamental for constatives is that at x 's expense, y can rely on p obtaining; this means that x is liable for compensating all damage which results for y from y 's erroneously relying on the fact that p .”

(von Savigny 1988, p. 51)

In this form, the requirement simply seems too strong—and indeed, if any utterance of a declarative incurred such a strong commitment, speakers would not utter them as frequently.

There are two possible responses: On the one hand, we might say that speakers actually take on this strong commitment, but it is rarely enforced. This seems to be a non-starter, because there are many cases where a sincere apology is all someone can ask for, even if one relied on someone's false assertion and had to suffer negative consequences because of it. On the other hand, we might broaden the notion of 'compensation for damages' so that something like an apology can serve as sufficient compensation. Or, perhaps, we simply can say that all one commits to is taking *partial* responsibility. This is what MacFarlane (2005) seems to have in mind:

“A plausible answer (though not the only one) is that part of what it is to make an assertion is to accept partial responsibility for the accuracy of what one says.”

(MacFarlane 2005, p. 319)

This version is not only weaker in that it only claims partial responsibility, it also does not tie the responsibility to negative effects of the hearer's action (though perhaps such negative effects can be generally assumed if the hearer's action was based on a false belief).

But it still seems too strong. Suppose, for example, both you and I observe the same evidence, which convinces us both (individually) that Mary was at some party. In our discussion of the evidence I utter (4.42), and you concur. Later it turns out that Mary in fact was not at the party at all, but rather she was studying. Given that we reasoned towards the conclusion that Mary was at the party together, it seems that you cannot hold me responsible for any harm that came your way due to your assumption that Mary was at the party.

(4.42) Mary was at the party.

Perhaps (4.42) is not an assertion in the sense that philosophers like MacFarlane (2011) are after, but it surely is a sincere utterance of a declarative. Perhaps we could say that by overtly agreeing (instead of just accepting) my utterance in (4.42), you absolved me from responsibility (cf. Gunlogson (2008)'s distinction of accepting and confirming declarative utterances). Or maybe we could say that the speaker is responsible only insofar as the addressee's reliance on the truth of the content of the utterance was due to the utterance.

But what, then, about the following case: I have heard (but you did not) from someone who both you and I usually trust as an authority about Mary's whereabouts (or about parties) that Mary was at the party; or perhaps I have other evidence that you and I would take to be conclusive of Mary's being at the party, even though you have not observed this evidence. When you wonder where Mary was yesterday, I utter (4.42). You accept my utterance, and get in a vicious fight with Mary, because you believe that she violated her duties by partying. It turns out that I was wrong. Am I liable, or to blame, for the adverse consequences to your relationship with poor Mary? I don't think I am. It seems that I would be well in my rights to express sympathy for your situation, but maintain that I only relayed information that I had excellent reason to think was true, and hence am entirely blameless.

4.5.4 Commitments to act as though one has a belief

Both (J) and (R) seem, in a way, too parochial: They capture intuitions about the obligations speakers take on in certain situations (philosophical discussion, and business deals, say), but they don't quite get at the heart of the matter: Declaratives are uttered in many contexts in which these effects do not seem to be present, or seem to be present only to a rather attenuated degree. Crucially, in such contexts, declaratives are often still very useful for expressing beliefs and to inform, and in these contexts, too, speakers are required to retract their utterances.

Condoravdi and Lauer (2011, 2011) propose a more general normative effect for utterances of declaratives. The idea is somewhat similar to the notion that

declaratives ‘express’ beliefs: A declarative utterance with content p commits its speaker to act as though he has the belief that p

(4.43) DECLARATIVE CONVENTION (normative)

A speaker who utters a declarative with content p thereby becomes committed to act as though he believes p to be true.

That is to say, if a speaker utters a declarative, he puts on himself an obligation to make only action choices that cohere with the belief that p (whether he actually has this belief or not). I will elaborate this proposal in the next section.

4.6 Commitments to act according to an attitude

The idea is this: Intentional actions are *chosen* by the agent performing them. Choosing an action means selecting it from a set of possible alternative actions (minimally: performing the action or not performing it), where this choice is (at least partially) determined by a form of *practical reasoning* about how to best satisfy one’s preferences, based on one’s beliefs.

I *choose* to grab my umbrella as I am leaving the house (rather than not grabbing it), because I believe that it is raining, or may be raining later in the day while I am out, and because I also have a preference against getting wet, and another belief that having an umbrella will enable me to stay dry in the rain. We may say that given this set of beliefs and preferences (together with some others, perhaps, such as that there will be no other way to procure an umbrella, and no other practical way to stay dry), practical reasoning tells me that it is *better* to grab my umbrella than not to. Idealizing a bit, we can say that choosing an action means performing the best action, or one of the best actions, from a set of alternatives, based on a given set of beliefs and preferences. *Acting as though one has a belief that p* , then just means choosing only those actions that are best with respect to a set of beliefs that includes p .

This will mean that a single commitment to a belief, by itself, will not completely

determine most action choices. If I sincerely utter (4.44), and thereby commit myself to act as though I believe that it will be raining, this is compatible with me not grabbing my umbrella. I may lack a preference for not getting wet, for example, or I may also believe that I will be in my office before it begins to rain, and that I won't have to leave it before it stops raining. And so on.

(4.44) It will be raining this afternoon.

This is desirable: If I only said (4.44), and you criticized me for not grabbing my umbrella, then informing you that I don't mind being in the rain without an umbrella, or that I don't plan to be outside all afternoon, etc. is a sufficient response.

However, my commitment to the belief that it will be raining will *constrain* my action choices, in particular together with other commitments I have. For I may be committed (or commit myself later) to other beliefs, and to having certain preferences. If I am committed to the belief that I will have to go outside in the afternoon, and I am committed to a preference for not getting wet, justifying my not taking the umbrella as being in accordance with my commitments will get harder. So the accumulation of commitments in dialogue can ultimately enable the audience to make rather reliable predictions about how I will act in the future (at least to the extent they take me to honor my commitments). Perhaps such predictions will never be quite *certain*—I might always be able add an unexpected belief or preference to subvert an action choice that otherwise would be mandated by my commitments—but the more commitments I incur, the more reliable my audience's predictions of behavior will become.

How does this conception—that declaratives commit the speaker to behave as though he had a belief—fare with respect to our three explanatory targets? The following subsections argue that it does well. Section 4.6.1 discusses (4.44c), Section 4.6.2 discusses (4.44a) and Section 4.6.3 discusses how various other normative effects, includes (4.44b) can be accounted for.

4.6.1 Declaratives with incompatible contents

It might be thought that the need for retraction in case of incompatible contents directly follows from the definition of the basic effect of declaratives—surely, committing to a belief that $\neg p$ is not a way to act as though one believes that p ? I am indeed on record with saying this (Lauer 2012, p. 395), but I now realize that it is not quite correct.

In Section 4.7, I will argue that it can be optimal for a speaker to assert p even if he does not believe it—but if that is the case, then committing to a belief that $\neg p$ *can* be a way to act as though one believed that p .

However, the proposed conception of declarative force does predict the need for retraction. While it may be rational to commit to a belief that one does not have, it is never rational to commit to two contradictory beliefs, which would amount to a commitment to act irrationally. So we explain the need for retraction not because committing to $\neg p$ is never a way to act as though one believed that p , but rather because committing to both p and $\neg p$ is never a sensible way to act (and so is not a sensible way to act as though one believes that p).

4.6.2 Declaratives express beliefs

This conception also directly explains why utterances of declaratives are (usually) reliable indicators of belief: While it may be, on occasion, advantageous for an agent to commit to a belief he does not actually have, any such a commitment will involve a *risk*. A speaker who does so risks committing himself to acting against his best interests.

For suppose a speaker believes that p , but commits himself to $\neg p$. At a later time, it turns out that, if p is true, that necessitates a certain action a , while if p is false, an alternative action a' would be optimal. The agent is committed to perform a' instead of a , even though his actual beliefs tell him that a would be better.

In such a case, of course, the speaker will not necessarily be forced to do a' . He might instead be able to rescind his commitment to $\neg p$. Or he simply chooses to not act in accordance with his commitment, and accept the blame for doing

so. But, by and large, it seems plausible that agents have incentive to abide by their commitments (if only to remain credible in the eyes of their peers) and also a weaker incentive to not rescind commitments all too frequently (for the same reason).

This means that, by and large, an audience has good reason to think that a speaker will commit to a belief that p only if he in fact believes that p . There will be exceptions—for example the speaker might not quite believe that p , but take it very likely that p obtains. Or it might be that the speaker believes that q , which is incompatible with p , but which necessitates all the same action choices as p in future situations he takes to be plausible. But in general, a speaker who utters declaratives that he believes to be true will be on the safe side, while a speaker who does not will engage in risky business.

For now, this informal sketch will suffice. Chapter 5 will be devoted to extending the system of dynamic pragmatics in such a way that we can model explicitly how an addressee infers that a speaker believes in the truth of his declarative utterances—based on the contextual assumption that the speaker prefers *not* to commit himself to beliefs he does not have.

4.6.3 Other normative consequences of declaratives

There is also some reason to think that the present conception has a role in explaining why some of the other proposals for normative consequences appeared promising—by explaining how, in special contexts, committing to act as though one has a certain belief commits one to certain courses of actions that can be described in other terms.

We have already seen that the present conception explains the need for retraction in case a speaker wants to make a new claim that is incompatible with his previous ones. This extends directly to cases in which the speaker *accepts* someone else's arguments as to why his previous utterances is false. This is what (W) mandated. Here is what MacFarlane says about this principle:

“Everyone should be able to agree that assertoric commitment includes

at least (W). Imagine someone saying: ‘I concede that what I asserted wasn’t true, but I stand by what I said anyway.’ We would have a very difficult time taking such a person seriously as an asserter. If she continued to manifest this kind of indifference to established truth, we would stop regarding the noises coming out of her mouth as assertions. We might continue to regard them as expressions of beliefs and other attitudes (just as we might regard a dog’s whining as an expression of a desire for food). We might even find them useful sources of information. But we would not regard them as commitments to truth.”

(MacFarlane 2005, p. 318)

On the present conception, it is clear why MacFarlane’s unfazed asserter would be viewed as doing something wrong: by saying **I concede that what I asserted wasn’t true, but I stand by what I said anyway**, not only does the speaker not rescind his previous assertion, he doubles down on it, while at the same time admitting that it was false, and thereby *committing* himself to its falsity. But that just means that he commits to inconsistent beliefs.¹⁴

Similarly, reconsider (J):

(J) Commitment to justify the assertion (provide grounds for its truth) if and when it is appropriately challenged.

¹⁴There still is an aspect of (W) as originally formulated that the present conception does not directly cover. The present perspective predicts that it is not possible (or at least not rational) to *accept* a conclusive argument that his prior assertion was false without withdrawing it. But (W) is stronger. It demands withdrawal whenever it is conclusively shown that one’s previous assertion was false, independent of whether one is willing to accept this publicly. Put differently, (W) implicitly mandates such public acceptance, while on its own, the present perspective does not: A speaker is well within his rights to not retract his previous assertion as long as he is willing to stick to his guns and refuse to publicly accept that he has been proven wrong.

I don’t think this is problematic: All we need to assume is that, at least in certain contexts, such as in rational debate, a speaker is independently required to accept such conclusive demonstrations of the falsity of his utterances. And I think that this is intuitively correct: A speaker who steadfastly refuses to back down from a previous assertion that has been conclusively proven wrong is not wrong because he fails to honor the commitments he made with his assertion. He is wrong because he fails to correctly respond to the proof that it was wrong.

I concurred with MacFarlane that it seems that this kind of commitment only applies in special cases, such as discussions among philosophers and other academics (or more general, *inquiry*). In the present context, we could explain the existence of the commitment in (J) by assuming that philosophers, etc., when engaged in discussions on their field of inquiry, are subject to an independent commitment to only believe (or commit to believe) things that they are able to defend against reasonable challenges. Together with the doxastic commitment created by an assertion, this would predict that a speaker, in such a context, is committed to utter only declaratives he is able to defend.

We can make similar assumptions about other cases where there are requirements on truth or epistemological warrant that are more stringent than in ordinary conversation. As an example, let us take witness testimony in a court of law: We can say that in this situation, a speaker is obligated to commit only to propositions he *knows* to be true (or perhaps: he is obligated to commit only to propositions that *are* true). That is, the commitment-based view, just as the expressive view, provides us with a 'hook' onto which further requirements, or moral, legal and social obligations can be hung. These requirements can pertain to what commitments an agent is allowed to take on.

4.7 The enduring commitments of loose talk: a case for the commitment-to-belief view

In the discussion of alternative views, I have occasionally remarked that, unlike the commitment-to-belief view articulated in the previous section, the alternative views by themselves predict the necessity of retraction only if this can be motivated on communicative grounds. In this section, I want to reprise the arguments made in Lauer (2012) to the effect that this provides an argument for the commitment-to-belief view.

In many cases (in particular typical instances of cooperative conversation which tend to be the focus of researchers on clause typing and speech act theory) almost

any of the alternative accounts can predict the need for retraction, given appropriate assumptions about cooperative behavior, essentially because all of these alternative accounts can, given appropriate assumptions, explain why declaratives are good for *informing* or *expressing beliefs*.

If a speaker now utters a declarative with content p that is incompatible with his previous assertions, he will often have reason to indicate that he has changed his mind (which he can do by retracting his previous assertion): He may want to make sure that his audience knows he had good reason for his previous utterance, he may want to alert them to the change in his doxastic state for other reasons. If he did not make this explicit, his audience might think he has forgotten that he made his previous utterance, and has forgotten the reasons why he made it, and might hence fail to believe his current utterance. That is, if he previously communicated $\neg p$, retracting this utterance will frequently be instrumental in successfully communicating p .

I am not certain that such a communication-based account of retractions could be spelled out in detail in such a way that it makes the correct predictions in all cases in which $\neg p$ is communicated prior to an instance where p was communicated. But I am also not certain that it could not be spelled out in this way. After all, we generally expect that belief revision is not the norm. We generally expect that speaker's beliefs persist, so maybe a case can be made that if a speaker changes his mind, he ought to (generally) indicate this, either because failing to do so would jeopardize his communicative goals, or because there is a social or moral norm requiring him to do so.

What I am certain about, however, is that this would not be enough. At best, such a communication-based account would predict that a speaker has to retract a previous assertion *whose content was believed by his audience*. But as we shall see, a speaker has to retract even utterances he knows have not been believed. What I have in mind are not utterances made in a situation in which the speaker was not believed because he was taken to be unreliable. Here, I want to focus on cases of perfectly cooperative communication.

4.7.1 Loose talk—the basic facts

Cooperative conversation may well involve utterances of declaratives that the speaker knows to be false, or at least, it would seem that way. Some examples of the phenomenon I have in mind are in (4.45).¹⁵

(4.45) Mary was here by three.

Fact: Mary arrived at 3:02.

(4.46) I am going to a conference in Berlin next week.

Fact: The conference is Potsdam, which abuts Berlin, but is not part of it.

(4.47) There were five hundred people at the rally.

Fact: There were exactly 489 people at the rally.

At least intuitively speaking, these sentences, in the indicated contexts, are *false* (though they ‘come close to being true’). And yet, there are many contexts in which a cooperative speaker who knows the indicated facts could use them. And such a speaker would not be at fault. This is what Lasersohn (1999) calls LOOSE TALK, or PRAGMATIC SLACK.

The philosophical literature provides a similar example, (4.48):

(4.48) France is hexagonal.

This example was introduced by Austin, who denied that, taken as a statement, (4.48) is true or false—it is simply ‘rough’:

“Suppose that we confront ‘France is hexagonal’ with the facts, in this case, I suppose, with France, is it true or false? Well, if you like, up to a point; of course I can see what you mean by saying that it is true

¹⁵(4.45) is a slight variant of an example that Lasersohn gives:

(i) Marry arrived at three.

I am no longer entirely convinced that Lasersohn was right in supposing that (i) involves ‘loose talk’ rather than truth-conditional vagueness. But his basic point is not affected if he has misclassified this example. (4.45) more clearly behaves in the way he has described.

for certain intents and purposes. It is good enough for a top-ranking general, perhaps, but not for a geographer. ‘Naturally it is pretty rough’, we should say, ‘and pretty good as a pretty rough statement’. But then someone says: ‘But is it true or is it false? I don’t mind whether it is rough or not; of course it’s rough, but it has to be true or false—it’s a statement, isn’t it?’ How can one answer this question, whether it is true or false that France is hexagonal? It is just rough, and that is the right and final answer to the question of the relation of ‘France is hexagonal’ to France. It is a rough description; it is not a true or a false one.”

(Austin 1962, p. 142)

Subsequent authors have taken a somewhat different view, saying that the sentence in (4.48) is not true or false *simpliciter*, but that it depends on the context whether it is—that is, they have argued that the sentence (and hence, presumably, the predicate **hexagonal**) is *vague*: Whether or not we take it as true depends on a contextually-given standard of precision. This is the view taken by Lewis (1979).

Lasersohn (1999), however, convincingly argues that this cannot quite be correct, or at least, it cannot be correct in all cases in which we feel that the speaker only ‘comes close enough to the truth’ with what he says. Because in many such cases, such as those cited above, there is good evidence that what the speaker said is *literally false*, while proper truth-conditional vagueness does not result in falsity. Thus we can conjoin a truly vague predicate like **bald** with a claim that there are exceptions.

- (4.49) a. Homer is bald: he has, like, three hairs left.
 b. Homer is bald, but he has a few hairs left.

This strongly indicates that **Homer is bald** can be *literally true* if Homer has some hairs left. By contrast, statements such as **Mary was here by three** do not allow for such conjunctions.

- (4.50) #Mary was here by three, but she did not arrive until a few minutes after three.

If **Mary was here by three** were truth-conditionally vague in the sense that **Homer is bald is**, the former sentence could be true even if she arrived shortly after three and (4.50) should be felicitous in such a context, but it is not. Lasersohn's conclusion is that loose talk is different from truth-conditional vagueness: It involves sentences that have context-independent truth-values, which yet can sometimes be 'blamelessly asserted' if they come close enough to the truth for practical purposes.

4.7.2 Loose talk in multi-sentence discourses

Turning our attention to multi-sentence (and multi-turn) discourses, we find that an assertion that is only loosely true cannot be felicitously followed up by a more precise one without further ado:¹⁶

- (4.51) (Yes,) Mary was here by three. # When she wasn't here five minutes after three, I got worried, but then she arrived in time.
- (4.52) A: I am going to this conference in Berlin.
B: Oh, where in Berlin is it held?
A: #In Potsdam(, which is just outside of Berlin).
- (4.53) A: Do we have enough coffee for the council meeting? How many people will be there?
B: Thirty.
A: Oh, great, then we'll have a quorum.
B: #No, because only 27 people will be there.

These discourses improve dramatically if we insert an expression that acknowledges the falsity of the previous assertion, such as **actually**:

¹⁶This short section is adapted from the one with the same title in Lauer (2012).

- (4.54) (Yes,) Mary was here by three. Well, actually, she wasn't here until five minutes after three, and I got worried, but then she arrived in time.
- (4.55) A: I am going to this conference in Berlin.
B: Oh, where in Berlin is it held?
A: Actually, it is in Potsdam(, which is just outside of Berlin).
- (4.56) A: Do we have enough coffee for the council meeting? How many people will be there?
B: Thirty.
A: Oh, great, then we'll have a quorum.
B: No, because (I was speaking loosely and) actually, only 27 people will be there.

At first glance, this may seem unsurprising: After all, we have just seen that there is good reason to believe that the first assertion in all these examples is literally false if the second one is true. So, of course, a speaker who wants to make the second assertion has to acknowledge this contradiction. Indeed, if the first assertion is appropriately hedged, no such acknowledgement is necessary:

- (4.57) (Yes,) Mary was here around three. When she wasn't here five minutes after three, I got worried, but then she arrived in time.
- (4.58) A: I am going to this conference in the Berlin area.
B: Oh, where is it held?
A: In Potsdam(, which is just outside of Berlin).
- (4.59) A: Do we have enough coffee for the council meeting? How many people will be there?
B: About thirty.
A: Oh, great, then we'll (likely) have a quorum.
B: No, because only 27 people will be there.

4.7.3 Loose talk in the commitment-to-belief account

Under the commitment-to-belief account of the sentential force of declaratives, we can construe cases of loose talk as instances where a speaker commits to a proposition that he knows to be false, or that he at least does not know to be true. Thus, in uttering (4.60) while knowing or suspecting that Mary arrived shortly after three, a speaker still commits himself to act as though she was here by 3:00:

(4.60) Mary was here by three.

Why would a speaker ever do this? He might have reason to do so, if he believes that the exact time of Mary's arrival is irrelevant to his future (linguistic and non-linguistic) action choices. That is, if he does not expect to be in a future decision situation in which there are two actions a and a' such that if Mary arrived at 3:00, a is optimal, but if she arrived slightly after 3:00, a' would be optimal. If there is no such future decision situation, then acting as though one believes that **Mary was here by three** is true is just *the same* as acting as though one believes that **Mary was here shortly after three** is true.¹⁷

Now suppose, contrary to the speaker's expectation, the conversational purposes shift so that more precision is required: Suddenly, it *does* make a difference whether Mary arrived at three or a couple of minutes after three. Given that the speaker has committed to a belief that she arrived at 3:00 sharp, we can explain why a speaker needs to acknowledge the falsity of his previous utterance.

(4.61) Actually, she wasn't here until a few minutes after three.

What **actually** does in this case, on the commitment analysis, is rescind the previous commitment taken on with (4.60), at the same time, (4.61) induces a new commitment to the effect that the speaker believes Mary arrived a few minutes

¹⁷An additional consideration such a speaker needs to take into account is whether his audience *knows* he is speaking loosely. If they take him to speak strictly—i.e., if they infer from (4.60) that Mary was here by 3:00, then a cooperative speaker may want to prevent this misunderstanding, and opt for a locution such as **around three**. He may, or he may not: If he does not expect that a couple of minutes of difference makes a difference to his audiences' future action, he may still opt to speak loosely.

after three.

By contrast, an utterance like (4.62), without **actually** or something that would do its job, would only create a new commitment, but leave the prior commitment intact. This is why (4.62) feels infelicitous after (4.60), even if the speaker was speaking loosely: He takes on incompatible commitments with his two utterances.

(4.62) She wasn't here until a few minutes after three.

It should be stressed that the commitments we are talking about here are, in most normal contexts, very 'cheap': The speaker can easily rescind them.¹⁸ But rescind them, he must, if he wants to make another assertion that is incompatible with them.

4.7.4 Communicative reasons for retraction

The prediction of the commitment-to-belief account is that the need for retraction is independent of the communicative goals of the speaker: It does not matter whether retraction is necessary to avoid a misunderstanding. Retraction is independently necessitated by the fact that declarative utterances give rise to commitments.

Let us reconsider the **Potsdam** example in a context in which a certain amount of slack is expected.¹⁹

(4.63) A: I am going to this conference in Berlin.

B: Where is it held?

A: # In Potsdam, which is just outside of Berlin.

It is noteworthy that *B*'s question does not indicate that he took *A* to speak strictly, and also that the continuation explicating that Potsdam is outside of Berlin does

¹⁸How 'cheap' rescission is will depend a lot on the context of utterance—if the context was such that precision was known to be necessary from the outset, rescission will be much harder than if initial demands were loose.

¹⁹It is helpful to assume that the conversation takes place somewhere far away from Berlin—in the US, say. If the interlocutors are close to, or in Berlin or Potsdam, it is implausible to assume that the speaker is speaking loosely, for reasons I cannot go into here.

not remove the sense of incoherence (if anything, it strengthens it). So, in this example, there is no *communicative reason* to retract: Even if the speaker does not retract, there is no danger that the addressee will misunderstand the speaker, and assume (for example) that Potsdam is part of Berlin.

To put this differently: Given that the speaker can be assumed to speak loosely, what *A* intends to convey with his initial utterance, and what *B* comes to believe if he takes *A* to be otherwise sincere and well-informed is something like the proposition expressed by (4.64).

(4.64) *A* is going to a conference in the Berlin area.

If all goes smoothly, (4.64) will be common belief after *A* uttered **I am going to this conference in Berlin**. *A* used this sentence to inform *B* that (4.64) is true, and he succeeded (let us assume). All parties to the conversation are in agreement that this is what happened. Why, then, can the speaker not act as though he had uttered (4.64), in which case no retraction is necessary?

(4.65) *A*: I am going to this conference in the Berlin area.
B: Where is it held?
A: In Potsdam, which is just outside of Berlin.

The commitment-based account explains why the speaker cannot just behave as if he made the 'loose' assertion in (4.65): Because he did not, and regardless of the fact that he communicated something weaker than the truth conditions of the sentence, he committed himself to act as though he believes the strict truth conditions.

Any account that instead tries to explain retraction facts solely in terms of the speakers' communicative goals will not be able to explain the contrast between (4.63) and (4.65) in a context in which slack is expected, because the two dialogues are the same in terms of what is communicated. It is only on the level of what is asserted that they differ.

To illustrate this another way: On the level of communicated content, a similar sequence involving a vague predicate like **bald** is completely parallel to (4.63):

- (4.66) A: Homer is bald.
B: Has he any hairs left?
A: Yeah, like three.

Again, *A*'s first utterance communicates that Homer is close to being completely bald, but not that he is completely hairless. *A* and *B* are aware of this. There is no communicative reason, then, to retract his first utterance when asserting that Homer has some hairs left. And in this case, because **Homer is bald** has vague truth-conditions, the commitment is to the weak truth conditions, too, and there is no need for retraction.

Loose talk provides us with an empirical argument to favor the commitment-based approach over alternatives that do not predict that utterances of declaratives induce commitment to the (strict) truth of their contents. Such accounts could conceivably explain the requirement for retraction where what is communicated is the strict truth conditions, but they cannot explain why retraction/acknowledgement is necessary when what is communicated is weaker than the truth-conditions of the asserted sentence.

Chapter 5

Action Choice and Commitment

At the end of Chapter 4, I proposed that declaratives give rise to *commitments to act as though one believes* the content of the declarative to be true and argued for a normative convention of the form in (5.1).

- (5.1) DECLARATIVE CONVENTION (normative) =(4.43)
A speaker who utters a declarative with content p thereby becomes committed to choose his actions as though he believes p to be true.

Among other things, this convention is supposed to help us explain why declaratives (in contrast to interrogatives and imperatives) are generally useful for conveying information, i.e., why it is that, when a speaker utters a declarative $\vdash \varphi$, this can lead to the addressee coming to believe that φ is true.

In Chapter 2, I developed a model in which we could show that an addressee will come to believe the content of an utterance φ if his belief state satisfies the conditions in (5.2).

- (5.2) *Contextual assumptions: Trusting addressee* =(2.34)
- a. 'Honest speaker'
 $B_{Ad,t} \models \text{utter}_t(Sp, \varphi) \Rightarrow \Box_{Sp,t}(\varphi)$
 - b. 'Informed speaker'
 $B_{Ad,t} \models \Box_{Sp,t}(\varphi) \Rightarrow \varphi$

The problem raised by the existence of multiple clause types is that (5.2a) should not hold for interrogatives and imperatives, but it should (usually) hold for declarative sentences. That is, we now want to use the convention in (5.1), together with plausible contextual assumptions, to *derive* (5.2a). The addressee-reasoning that we want to capture is given informally in (5.3).

- (5.3)
- a. [Convention]
When a speaker utters $\vdash \varphi$, he becomes committed to believe φ .
 - b. [Contextual assumption]
The speaker does not want to be committed to believe φ unless he actually believes it.
 - c. [From (a) and (b)]
The speaker will decide to utter $\vdash \varphi$ only if he believes that φ .
 - d. [Belief about actions]
Since utterances are intentional actions, the speaker will utter $\vdash \varphi$ only if he decides to do so.
 - e. [From (c) and (d)]
The speaker will utter $\vdash \varphi$ only if he believes that φ .

This reasoning is rather straightforward on the surface, but there is much hidden complexity. Besides reasoning about *beliefs* and *commitments*, it involves a premise about *desires* or *preferences*, viz., (5.3b), and a conclusion about what the speaker will *decide* to do, given his preferences, viz., (5.3c), as well as a seemingly innocent assumption about intentional actions, viz., (5.3d).

If we want to capture such reasoning—if we want to explain how a convention like (5.1) enables speakers to communicate that they believe in a proposition, our system of dynamic pragmatics needs to be able to capture reasoning about how speakers decide on (utterance) actions, based on their preferences and beliefs.

An additional benefit of representing action choice is that this will enable us to be more explicit about the nature of normative conventions like (5.1). The convention says that the speaker becomes committed to choose his actions in a certain way. In order to give this notion content, we need to be able to talk about action choice.

5.1 Modeling action choice

At its most abstract, deciding what to do can be construed as selecting an action from a set of alternatives (minimally, doing a certain act or not doing it). Which alternative an agent selects depends—besides the set of available action alternatives—on what outcomes he wants to achieve, and what he believes about the consequences of each of the actions. That is, we can abstractly characterize a decision procedure as in (5.4).

- (5.4) A *decision procedure* is a function Opt that takes three arguments B , P and A ; where B and P are suitable representations of beliefs and preferences, and A is a set of possible actions, such that $\text{Opt}(B, P, A) \subseteq A$.

This way of construing action choice is, of course, not a novel idea at all. It is a familiar paradigm in many disciplines that are modeling how people choose their actions. Most notably, this conception of action choice is the one of decision and game theory. The most dominant approach there is to assume that beliefs are specified as PROBABILITY DISTRIBUTIONS, preferences are specified by means of UTILITY ASSIGNMENTS, and Opt is defined in terms of the EXPECTED UTILITY of the actions. In other words, Opt returns the subset of A which contains only those actions that maximize expected utility.

The central concern of these approaches is *choice under uncertainty* about the consequences of the actions: Agents are taken to have preferences over fully-specified outcomes, but are uncertain as to which action results in which outcome. Expected utility theory and its alternatives are theories about how an agent makes such a choice, given his probabilistic beliefs about the world.

However, the usefulness of the conception of action choice in (5.4) is not exhausted by these concerns. Indeed, in this dissertation, the conception of a decision procedure that selects among actions, given certain input beliefs and preferences will be very central indeed, while uncertainty about the outcomes of actions will not play a central role at all.

Given a known Opt -function, we can use information about its output to draw inferences about its inputs: Suppose that B_1 and P_1 are unknown, and for some set of actions A , we are told that one of its members a is such that $a \in \text{Opt}(B_1, P_1, A)$. Then, using our knowledge of Opt , we can infer constraints on what B_1 and P_1 must be like. Of course, we will never be able to fully know either, but especially if we had some partial antecedent information about both, then learning that $a \in \text{Opt}(B_1, P_1, A)$ can allow us to infer rather concrete things about B_1 and P_1 .

The reasoning in (5.3) is of just this kind: By observing an utterance, the addressee can conclude that this utterance satisfied the speaker's preferences better than the alternatives, given his beliefs; and so the addressee can 'reason backwards' to a conclusion about the speaker's beliefs. We can bring this out by reformulating (5.3). Let B_{Sp} and P_{Sp} be the speaker's beliefs and preferences, and let A_{Sp} be any set that contains $\text{utter}(Sp, \varphi)$.

- (5.5)
- a. [Knowledge of convention]
According to B_{Sp} : When Sp utters $\vdash \varphi$, he becomes committed to believe φ .
 - b. [Contextual assumption]
 P_{Sp} specifies that the speaker will not commit himself to φ unless B_{Sp} supports φ .
 - c. [From (a) and (b) + knowledge of Opt]
 $\text{utter}(Sp, \varphi) \in \text{Opt}(B_{Sp}, P_{Sp}, A_{Sp})$ only if B_{Sp} supports φ .
 - d. [Belief about actions]
For any $a \in A$: a happens only if $a \in \text{Opt}(B_{Sp}, P_{Sp}, A_{Sp})$.
 - e. [Observation]
 $\text{utter}(Sp, \varphi)$ happened.
 - f. [From (d) and (e)]
 $\text{utter}(Sp, \varphi) \in \text{Opt}(B_{Sp}, P_{Sp}, A_{Sp})$
 - g. [From (c) and (f)]
 B_{Sp} supports φ .

The system \mathcal{P}_{Sen} already contains a representation of beliefs, allowing us to model premise (5.5a), as well as the conclusion (5.5g), and it allows us to talk about events occurring, as in premise (5.5e). In order to capture the reasoning in (5.5) fully, we minimally need to add a representation of preferences (so that we can represent premises like (5.5b)) and we need to specify how Opt relates actions to beliefs and preferences (so as to represent a premise like (5.5c)).

Section 5.2 introduces preferences, Section 5.2.2 a working definition of Opt . Before, I will show how to implement (5.5d), the assumption that speakers only perform actions that are selected by a given Opt . This involves some complications, which will lead us to revise one of the constraints we have put on beliefs in Chapter 2.

5.1.1 Actions and agents

As a first step, we need to amend our model to specify which events are actions, and who the agent of a given event is (i.e., who decides whether an action occurs). Let Ag be the set of agents, then:

(5.6) The models for \mathcal{P}_{Sen} contain a partial function $Agt : W \times T \mapsto Ag$.

Agt specifies the unique agent (if any) who acts at t in w . Given that we have restricted our models such that only one event can happen at any given time, only one agent can act at any given time, hence it makes sense to also require that there is a unique agent who decides on actions. We also define the function $Act : W \times T \mapsto \mathbb{E}$, which specifies action alternatives for the agent.

(5.7) For all $w, t : Act(w, t) = \{ev \mid \exists v : w \approx_t v \ \& \ \mathbf{Hap}_v(t, t + 1) = ev\}$
if $Agt(w, t)$ is defined, undefined otherwise.

Thus if $Agt(w, t) = i$, then agent i decides which of the possible futures that diverge at time t becomes actual. The alternative actions available to him are all events that can happen just after t . We don't require that $Agt(w, t)$ is defined for all w, t to allow for events that are not actions.

5.1.2 Beliefs about action choices

Given a fixed Opt , we now simply impose a generalized version of premise (5.5d) as a constraint on our models. The constraint requires that all agents believe each other to only perform actions that cohere with Opt . This is an idealization, even with utterance actions. Sometimes a speaker may speak involuntarily, in a way that does not cohere with his beliefs and preferences, and his audience may well be aware of this fact. But pragmatic reasoning, or at least pragmatic reasoning of the Gricean kind, is suspended in such a situation and hence, in situations that we are interested in as pragmaticists, this assumption is warranted.

(5.8) **Belief in optimal actions constraint**

For all $w, v, t, t', i \neq i'$ such that $wR_{i,t}v$ and $\text{Agt}(v, t') = i'$:

$$\text{Hap}_v(t', t' + 1) \in \text{Opt}(B_{i',t',v}, P_{i',t',v}, \text{Act}(v, t'))$$

where $P_{i,tvw}$ is a representation of the preferences of i' at v, t .

This constraint straightforwardly captures the assumption that agents believe each other to choose their actions via Opt (based on their beliefs and preferences).

Given that we want to implement this constraint, however, we need to slightly relax one of the other constraints we imposed on our models in Chapter 2: the constraint **No fore-knowledge**, repeated below, which requires that the belief-relations at any given time t do not distinguish between worlds that have not diverged at t : If an agent takes a world to be possible, he must also take all its possible futures to be possible.

(5.9) **NO FORE-BELIEF constraint** =(2.16)

If $v_1 \approx_t v_2$, then $wR_{i,t}v_1$ iff $wR_{i,t}v_2$

It may not be immediately obvious that (5.8) and (5.9) are in conflict. On the face of things, combining the two simply seems to have the result that an agent can take a world v to be possible at t only if all possible futures of v are such that only optimal

actions are chosen at times after t .

That, precisely, is the problem. A world v which only has futures in which optimal actions are chosen at all times is a world in which no non-trivial action choices occur.¹ Here is why: The actions available to an agent at t' are determined by the possible futures at t' : For each possible action, there is at least one such possible future in which the action occurs. This includes the actions that are *non-optimal*, which will exist in all non-trivial choice situations. But then, every world in which a non-trivial action choice is made will have possible futures at which non-optimal actions are performed.

This is not an accident, but a necessity: By hypothesis, an agent chooses between his actions by comparing what *would* be the case if he performed them. But that means that even when he acts optimally, there must be 'counterfactual' possibilities in which he does not, because otherwise we cannot define what it means to act optimally). The relevance of such counterfactual possibilities in modeling action choice has been recognized in the epistemic approach to game theory, in particular in connection with the justification of backward-induction in games of perfect information (Aumann 1995, Aumann 1996, Stalnaker 1996).

What this means that if an agent i believes about another agent that he will act optimally in the future, then i has a fore-belief: On the one hand, he takes the world to be such that the agent could perform both optimal and non-optimal actions (for else, we cannot say that the agent will *choose* an action). On the other hand, i believes that some of these actions (the non-optimal ones) will not occur. Seen this way, beliefs about future action choices intrinsically *are* fore-beliefs.

However, we do not want to give up on the constraint in (5.9) altogether. We still require it to hold with respect to non-action events. The following sequence of definitions defines the weaker version of No FORE-BELIEF that ensures this.

(5.10) For two worlds w_1, w_2 such that for some $t : w_1 \approx_t w_2$, the *point of divergence* $div(w_1, w_2)$ is the unique time t' such that $w_1 \approx_{t'} w_2$ and $w_1 \not\approx_{t'+1} w_2$.

¹Formally, such a world v must be such that at all future times t' and all worlds $v' \approx_t v$: If $Act(v', t')$ is defined, then $Opt(B_{Ag(t'), t', v'}, P_{Ag(t'), t', v'}, Act(v', t')) = Act(v', t')$, i.e., it must be that, at all future decision points, all alternative actions are equally optimal.

- (5.11) Two worlds w_1 and w_2 are *external historical alternatives* at t , written $w_1 \approx_t^e w_2$ iff $w_1 \approx_t w_2$ and $\text{Agt}(w_1, \text{div}(w_1, w_2))$ is undefined.
- (5.12) **No fore-knowledge about non-actions** constraint
If $v_1 \approx_t^e v_2$, then $wR_{i,t}v_1$ iff $wR_{i,t}v_2$, for all i .

5.2 Modeling Preferences

We now have extended our models so that they contain the assumption that agents take each other to choose their actions, based on their beliefs and preferences. In terms of the reasoning that we want to capture, we are now able to represent the steps marked with a ✓ below.

- (5.13) a. [Knowledge of convention]
According to B_{Sp} : When Sp utters $\vdash \varphi$, he becomes committed to believe φ .
- b. [Contextual assumption]
 P_{Sp} specifies that the speaker will not commit himself to φ unless B_{Sp} supports φ .
- c. [From (a) and (b) + knowledge of Opt]
 $\text{utter}(Sp, \varphi) \in \text{Opt}(B_{Sp}, P_{Sp}, A_{Sp})$ only if B_{Sp} supports φ .
- d. [Belief about actions]
✓ For any $a \in A$: a happens only if $a \in \text{Opt}(B_{Sp}, P_{Sp}, A_{Sp})$.
- e. [Observation]
✓ $\text{utter}(Sp, \varphi)$ happened.
- f. [From (d) and (e)]
✓ $\text{utter}(Sp, \varphi) \in \text{Opt}(B_{Sp}, P_{Sp}, A_{Sp})$
- g. [From (c) and (f)]
✓ B_{Sp} supports φ .

All that is missing (besides a way to talk about *commitments*, which is the topic of the second part of this chapter) is a suitable representation of preferences (which

we have presupposed anyway), so as to represent (5.13b), as well as as a characterization of Opt that allows us to conclude (5.13c). This is what we will do in the present section. We begin with the representation of preferences in Section 5.2.1. Thereafter, in Section 5.2.2, I discuss possibilities for Opt that support the inference in (5.13c).

5.2.1 Preference structures

I will represent the preferences of an agent as a set of *propositions*, which are ordered in terms of their *importance*. We can think of these propositions as capturing *outcome preferences*, i.e., as properties of outcomes that the agent prefers.² In other words, if the proposition p is a preference of an agent, this means that he wants to see p realized, all else being equal.

Now, of course, an agent typically has many different kinds of preferential attitudes: desires, inclinations, appetites, personal moral codes, and so on. I assume, however, that all these diverse preferential attitudes can be integrated into one over-arching attitude that an agent uses to guide his action choices. This assumption of an ‘overall’ preferential attitude is again familiar from decision and game theory: There, too, an agent’s action-relevant preferences are typically represented by a single utility assignment, which is viewed as the composite of all preferential factors guiding the agent’s action choice. I call these ‘overall’ preferences **EFFECTIVE PREFERENCES** (a term introduced in Condoravdi and Lauer 2011), to highlight the fact that these are the preferences that are relevant for the agent’s action choices.

(5.14) A *preference structure* is a pair $\langle P, \leq \rangle$, where P is a set of propositions and \leq is a binary relation on P that is reflexive, transitive and total.³

²In Chapter 9, when I apply the model of preference and action choice to conversational implicatures, it will be necessary to represent a second kind of preference, *action preferences*, for which a propositional format is not well-suited.

³A binary relation R on a set P is **REFLEXIVE** if pRp for all $p \in P$, it is **TRANSITIVE** if for all $p_1, p_2, p_3 \in P$, p_1Rp_2 and p_2Rp_3 imply p_1Rp_3 and it is **TOTAL** if for all $p_1, p_2 \in P$, either p_1Rp_2 or p_2Rp_1 (and possibly both).

I take it that reflexivity and transitivity are intuitively necessary properties for any relation that captures an ‘importance ranking’ which allows for ties. Totality might not be as obviously adequate as a requirement—an agent might be genuinely undecided which of two preferences is more important, without necessarily considering the two preferences *equally* important.⁴ Requiring totality has the convenient consequence that a preference structure decomposes into equivalence classes that are linearly ordered by \leq . As it will occasionally be useful to refer to these equivalence classes, I introduce some notational conventions to talk about them:

- (5.15) Notational conventions: Given a preference structure $\langle P, \leq \rangle$,
- a. $p < q$ iff $p \leq q$ and $q \not\leq p$
 - b. $p =_{\leq} q$ iff $p \leq q$ and $q \leq p$
 - c. $[p]_{\leq} = \{q \mid p =_{\leq} q\}$
 - d. \equiv_{\leq} is the set $\{[p] \mid p \in P\}$

Preference structures in themselves can be used to represent any kind of preferential attitude. Effective preferences have to satisfy a number of additional rationality constraints, in particular they must be both *realistic* and *consistent*, relative to the agents’ beliefs.

(5.16) **Realistic preference structures**

An preference structure $\langle P, \leq \rangle$ is *realistic*, relative to an information state B iff for all $p \in P$: $p \cap B \neq \emptyset$.

(5.17) **Consistent preference structures** (Condoravdi and Lauer (2012))

A preference structure $\langle P, \leq \rangle$ is *consistent* with respect to an information state B iff for any $X \subseteq P$, if $B \cap \bigcap X = \emptyset$, there are $p, q \in X$ such that $p < q$.

The definition of consistency is a generalization of the definition in Condoravdi and Lauer (2011), which perhaps is easier to grasp intuitively:

⁴This seems to be possible, in particular, for non-effective preferences, such as mere desires or inclinations. If so, it would be more adequate to impose totality as a rationality constraint on *effective* preferences, on a par with the realism and consistency constraints introduced below.

(5.18) **Consistent preference structures** (simplified version, Condoravdi and Lauer (2011))

A preference structure $\langle P, \leq \rangle$ is *consistent* with respect to an information state B iff for any $p, q \in P$, if $B \cap p \cap q = \emptyset$, then $p < q$.

(5.18) says that a preference structure is consistent (relative to an information state B) if every two preferences that are known (in B) to be incompatible are strictly ranked. The underlying intuition is that if an agent believes that two of his preferences are incompatible, in order to act, he has to decide which of the two is more important to him. (5.17) generalizes this condition to ensure strict ranking of preferences also in case more than two propositions are known to be jointly incompatible, even if they are pairwise compatible. Consistency in the sense of (5.17) entails consistency in the sense of (5.18).

With these notions in place, we amend the models for \mathcal{P}_{Sen} again:

(5.19) Models for \mathcal{P}_{Sen} contain a function EP such for all $i \in Ag, w \in W, t \in T$, $EP_i(w, t)$ is a preference structure that is realistic and consistent with respect to $B_{i,t,w}$.

Given that effective preferences are those preferences that guide action choice, both constraints on them are intuitively appropriate. An agent should not strive to make true a proposition he knows is unattainable; and if he has two preferences he knows to be inconsistent (an appetite for ice cream, say, as well as a desire to lose weight), he'd better decide which of the two preferences is more important to him before deciding to act.

Ranked preferences as *ceteris paribus* preferences

Employing ranked preferences (instead of just a flat set of propositions) has various advantages which will become apparent in what is to follow. The most important advantage is that it lets us conceive of the propositions in EP as *ceteris paribus* preferences: If $p \in EP$, that does not necessarily mean that the agent prefers p over $\neg p$ tout court, but rather only if 'everything else is equal'. Hansson (1996)

articulates well how central such preferences are to our everyday reasoning:

“When discussing with my my wife what table to buy for our living room, I said: ‘A round table is better than a square one.’ By this I did not mean that, irrespective of their other properties, any round table is better than any square table. Rather, I meant that any round table is better (for our living room) than any square-shaped table that does not differ significantly in its other characteristics, such as height, wood, finishing, price, etc. This is preference *ceteris paribus* or ‘everything else being equal’. Most of the preferences that we express or act upon seem to be of this type.”

(Hansson 1996, p. 307)

Such *ceteris paribus* preferences play a crucial role in pragmatic reasoning, in particular in the derivation of implicatures, which will be discussed in Chapter 9. To anticipate, it is often necessary to assume that speakers prefer uttering certain forms over uttering others. An often used example is a preference for forms that are short and simple over those that are longer and more complex (cf. Grice’s (1975) MAXIM OF BREVITY).

While it is intuitive to assume such a preference in general, in usual cases such an assumption makes sense only if the preference is *ceteris paribus*: It should be subordinate to the speaker’s communicative goals, for example. If a speaker desires to convey a certain amount of information (say, $p \wedge q$) that cannot be conveyed by using a shorter form (say, ‘ p ’ or ‘ q ’ instead of ‘ $p \wedge q$ ’), then we usually would not expect the speaker to give up on his communicative goal in order to satisfy his preference for a shorter form—that is, his preference for shorter forms is only *ceteris paribus*: In choosing between two forms, the speaker will opt for the shorter form only if both forms satisfy his other goals equally. The order \leq is intended to capture this—intuitively, a preference $p \in P$ is only to be taken into account if the preferences that strictly outrank p do not decide between two actions.

Interaction between beliefs and preferences

Effective preferences are those that an agent uses to decide for actions. It hence makes sense to assume that agents are *aware* of their effective preferences. We ensure this by requiring:

(5.20) **Preference introspection constraint**

$$p \in EP_a(w, t) \text{ iff for all } v \in B_{i,t,w} : p \in EP_i(v, t)$$

(5.20) embodies an idealization, in particular together with the assumption that agents take each other to always act optimally. In actual fact, an agent's decisions can be influenced by preferential attitudes he is unaware of. And yet, in such a case, an agent will generally *assume* that he is aware of the factors driving his action choices. In order to implement this, we could weaken (5.20) to only apply to worlds contained in the agent's belief state, but not necessarily the actual world. However, since preferences that the agent is unaware of will not play a role in the cases I discuss in this dissertation, I will assume the simpler (5.20) throughout.

We will also introduce an operator ep that allows reasoning about effective preferences in the pragmatic language, but before we can do so, we will have to specify how preferences (and beliefs) relate to action choice.

5.2.2 A working definition for Opt

The framework now can represent beliefs and preferences. All that is needed to reason about action choice is a specification of Opt . This operator is, of course, central to the whole enterprise. Yet, I will not spend much time in defending the choice of Opt made here. The reason for this is that the precise nature of the decision procedure is a very complex topic, which depends, in no small part, on considerations beyond the scope of pragmatic theory. Ultimately, we want a definition of Opt that is useful to model action choice in general, with utterance choice just being a special case. But, of course, investigations of language use will not tell us anything about how non-linguistic actions are chosen. A full, and generally defensible, specification of Opt can only emerge from the joint work of

pragmaticists and researchers in other disciplines concerned with action choice, such as psychology, behavioral economics, or the philosophy of action.

Luckily, however, we do not *need* to have a generally-applicable definition of Opt in order to investigate pragmatic reasoning. We can make do with imposing some constraints on how Opt functions.

In order to state these constraints, it is useful to conceive of $\text{Opt}(B, P, A)$ as being defined in terms of a ‘betterness’ relation $<_{B,P}$ over the actions in A , such that $a < a'$ iff a' is strictly better to fulfill the preferences in P , given the beliefs in B . Given that it is supposed to be a strict betterness relation, we require that $<_{B,P}$ is irreflexive and transitive (hence anti-symmetric). Then Opt is simply defined as:⁵

$$(5.21) \quad \text{Opt}(B, P, A) := \{a \in A \mid \text{for no } a' \in A : a <_{B,P} a'\}$$

if $B \neq \emptyset$ and P is realistic and consistent given B . Else undefined.

The only substantive constraint on admissible $<_{B,P}$ orderings I will impose here is that it should respect the ‘importance’ ranking \leq on P : As indicated above, the idea is that a lower-ranked preference is taken into account only if the higher-ranked preferences don’t make a decision. (5.22) captures this idea.

(5.22) **Lexicographicness**

If $a <_{B,P} a'$ and $P \subseteq Q$ such that for all $p \in P, q \in Q \setminus P : q < p$,
then $a <_{B,Q} a'$.

Beyond this, there are only very weak assumptions that should be uncontroversial, given the interpretation of the propositions in P as properties of outcomes that the agent prefers:

- (5.23)
- a. For no $a, a', B : a <_{B, \emptyset} a'$.
 - b. If for all $p \in P : B[a] \subseteq p$, then for no $a' : a <_{B,P} a'$.
 - c. If for all $p \in P : B[a'] \subseteq \neg p$, then for no $a : a <_{B,P} a'$.

⁵Again, defining Opt via a relation on actions is very much in keeping with the conception of action choice in decision and game theory: There, the ‘betterness’ ranking is defined in terms of expected utility, and the prediction of the theory is that agents will (or should) choose one of the actions that yields the highest expected utility value.

(5.23a) says that an agent that does not have any preferences should be indifferent between any two actions. (5.23b) requires that if an action is certain to satisfy all preferences of the agent, it should be optimal and (5.23c), conversely, says that an action that is certain to *not* satisfy any of the preferences of an agent should not be strictly preferred over any other action.

Besides these requirements, the shape of $<$ (and hence, the nature of Opt) largely depends on two questions: (i) How is uncertainty about the consequences of actions treated? and (ii) How are preferences that are unranked with respect to each other treated?

For the sake of concreteness, I will define a particular instantiation of $<_{B,P}$ that embodies very simple answers to both questions. This is defensible largely because the cases we will consider in the remainder of the thesis will be independent from the correct answer to (i) and (ii). The definition of $<_{B,P}$ offered below hence should be seen as nothing more of an example implementation that is useful for working through example cases: All that I require in general are the constraints given above.

(5.24) Given a belief state B and a preference structure $\langle P, \leq \rangle$; and $E \in \equiv_{\leq}$ let

$$a \leq_E a' \text{ iff } \{p \in E \mid B[a] \subseteq p\} \subseteq \{p \in E \mid B[a'] \subseteq p\}$$

with \approx_E and $<_E$ defined in the obvious way in terms of \leq_E .

(5.25) **Betterness of actions** (exemplary version used in the rest of this thesis)

$$a <_{B,P} a' \text{ iff for some } E \in \equiv_{\leq} : a <_E a' \text{ and for all } E' > E : a \approx_{E'} a'$$

With respect to question (i) above, this definition simply assumes that all that matters is whether or not an agent is certain that an action realizes a given preference (i.e., whether or not $B[a] \subseteq p$), and hence is a lexicographic variant of a ‘maximin’ rule (i.e., it optimizes the ‘worst case’). This is sufficient for the cases discussed in the remainder of this dissertation, as substantial uncertainty about the consequences of actions will not play a role.

With respect to question (ii), the definition makes a very weak claim: An action

a is strictly better than another action a' iff a satisfies all the preferences that a' satisfies, and also satisfies at least one preference that a' does not satisfy. At least as long as we restrict attention to cases in which the agent is certain which preference any given action satisfies, this seems like a minimal requirement on a 'betterness' relation for agents.⁶

5.2.3 Reasoning about preferences

As a final step, we now introduce an ep operator into the pragmatic language \mathcal{P}_{Sen} , so that we are able to reason about effective preferences.

The intended interpretation for $ep_{i,t}(\phi)$ is that at time t , agent i 's effective preference structure is such that he acts as if ϕ were a maximal preference of his. To implement this, we first define an operator $+$ which adds a new preference to an existing preference structure as a maximal element:

$$(5.26) \quad \langle P, \leq \rangle + \phi = \left\langle P \cup \{\phi\}, \leq \cup \left\{ \langle \phi, p \rangle \mid p \in P \right\} \cup \left\{ \langle p, \phi \rangle \mid p \in \max(P, \leq) \right\} \right\rangle$$

Secondly, we define an equivalence relation over preference structures, as follows:⁷

$$(5.27) \quad EP \sim_{i,t,w} EP' \text{ iff} \\ \forall v \in B_{i,t,w}, t' \geq t : \text{if } Agt(w, t) = a, \text{ then} \\ \text{Opt}(B_{i,t,w}, EP, Act(w, t')) = \text{Opt}(B_{i,t,w}, EP', Act(w, t'))$$

That is, two preference structures are $\sim_{i,t,w}$ -equivalent iff at all future decision situations which the agent expects to face, EP and EP' determine the same set of optimal actions.⁸ The idea is that two preference structures are equivalent if, given

⁶This definition is essentially the same as the definition of rationality used by Aumann (1995, 1996) and related work in game theory when there is no uncertainty about outcomes (in particular, in games of 'perfect information'): "a player i is rational if and only if it is not the case that he knows that he would be able to do better" (Aumann 1996, p. 138). Of course, for Aumann, 'do better' is defined as maximizing a utility value, in which case this definition comes out as a consequence of expected utility maximization: "This is weaker than (i.e., implied by) expected utility maximization; if an agent is maximizing expected utility, surely he cannot have an option that he knows, with certainty, will yield a preferred outcome." (Aumann 1995, p. 9, n. 4)

⁷Recall that because the R_a relations are transitive and Euclidean, we have $\forall v \in B_{i,t,w} : B_{i,t,w} = B_{i,t,w}$.

⁸In full generality, we probably want to restrict the set of future decision situations that are

what the agent believes, it does not make a difference which one he uses to choose his actions.

With these notions in place, we can define ‘preference-support’:

$$(5.28) \quad EP \vdash_{i,t,w} \phi \text{ iff } EP + \phi \sim_{i,t,w} EP \text{ and } EP + W \setminus \phi \not\sim_{i,t,w} EP$$

EP supports ϕ if adding ϕ as a maximal element to EP would not change the agent’s decision in any situation he expects to face. The second conjunct in the definition is necessary to exclude ‘spurious’ support—that is, if an agent expects that neither a ϕ preference nor a not- ϕ preference would make a difference for his future decisions, we do not want to say that his preferences support ϕ —e.g., we do not want to say that an agent prefers it to rain tomorrow in virtue of the fact that he does not have control over the weather.

With this, we can state the interpretation clause for ep:

$$(5.29) \quad w \models \text{ep}_{i,t}(\phi) \text{ iff } EP_i(w, t) \vdash_{i,t,w} \phi \quad (\text{final})$$

This ensures that if p is a maximal element of $EP_i(w, t)$, then $w \models \text{ep}_{i,t}(p)$, which is how Condoravdi and Lauer (2011) defined the ep operator. This was sufficient there, as we were mostly interested in talking about the *public* version of effective preferences, which only have maximal elements (cf. Section 5.3 below).

The definition in (5.29) is more general, as it also allows $\text{ep}(p)$ to be true if p is a non-maximal element, but the agent does not expect to face a decision situation in which p is defeated by a higher-ranked preference.

5.2.4 Summary

With the representation of preferences, and a definition of Opt, we are now able to formalize most parts of the reasoning schema we started out with—except for the parts pertaining to commitment:

taken into account somewhat. It may be that an agent only expects to face a decision situation after learning some new pertinent fact. In that case, this decision situation should not be taken into account.

- (5.30) a. [Knowledge of convention]
 $B_{Sp}[\text{utter}(Sp, \varphi)] \models Sp$ is committed to φ
- b. [Contextual assumption]
 $ep_{Sp}(Sp \text{ is committed to } \varphi \Rightarrow B_{Sp} \models \varphi)$
- c. [From (a) and (b) + knowledge of Opt]
 $\text{utter}(Sp, \varphi) \in \text{Opt}(B_{Sp}, P_{Sp}, A_{Sp}) \Rightarrow B_{Sp} \models \varphi$
- d. [Belief about actions]
 For any $a \in A$: a happens only if $a \in \text{Opt}(B_{Sp}, P_{Sp}, A_{Sp})$.
- e. [Observation]
 $\text{utter}(Sp, \varphi)$
- f. [From (c) and (d) and (e)]
 $B_{Sp} \models \varphi$

More generally, this reasoning schema captures the following inference: If Sp believes that an action a has consequence c , and he prefers to avoid c unless that c' obtains, then from observing a , we can conclude that he believes c' obtains.

The relevant consequence c here is ' Sp becomes committed to believing φ ' and the condition is 'The speaker believes φ '. The rest of this chapter lays out how to capture the notion of commitment in the system of dynamic pragmatics.

5.3 Modeling commitment

The notion of commitment I am aiming at here may not completely match the everyday use of the word **commitment**, and it certainly is at variance with ways in which the term has been used in linguistics and philosophy. Before I go ahead with saying how this notion should be understood formally, I want to briefly lay out some basic properties I take commitments, in the intended sense, to have.

Commitments are public I take commitments to be always inter-personal. That is, a commitment is always made towards an agent, or a group or community of agents. I take this to be a fundamental part of our notion of 'commitment': One

cannot take on a commitment unless there is at least a witness to whom one makes it. Of course, we sometimes talk of private resolutions as if they were commitments: If an agent privately resolves to exercise at least three times a week, we may say that he 'makes a commitment' to exercising that often. I take such uses to be derived from the inter-personal sense.

Commitments are *commitments to act* I also take commitments to be necessarily connected to *action*: A commitment is kept by making the right action choices; it is violated by making the wrong ones. That means I take locutions such as '*a* is committed to believe / the belief that *p*' to be short for '*a* is committed to act as though he believes *p*'. Consequently, if an agent is 'committed to the belief that *p*', he does not violate this commitment merely by failing to believe that *p*. Instead, he can be in perfect compliance with his commitment, as long as his action choices are consistent with (his other commitments and) the belief that *p*.⁹

Conceiving of commitments in this way quite severely limits the things one can be committed to: An agent can only be committed to things that influence his action choices, i.e., an agent can only be committed to mental attitudes (strictly speaking, to act as though he had a mental attitude), in particular beliefs and preferences. Saying that an agent 'is committed to *p*', where *p* is a proposition about an agent-external fact makes sense only as an abbreviation for 'is committed to (act though he has) the belief that *p*' or, possibly, 'is committed to (act as though he has) a preference for *p*'.¹⁰

Commitments survive the present discourse The notion of commitment employed here is more general than the 'mere discourse commitments' in theories like that of Hamblin (1971) and the work of Gunlogson (2003, 2008). These authors

⁹This will be instrumental in explaining why a speaker would be willing to ever engage in 'loose talk', e.g., saying that France is hexagonal even though he clearly believes it is not. In the theory developed here, such an utterance will commit the speaker to the belief that France is hexagonal—and, by and large, we will predict that a speaker will only utter the sentence if he believes the non-hexagonality of France will not make a difference to his future action choices.

¹⁰Both kinds of abbreviatory uses are common, e.g., **John is committed to Toulouse being the capital of France** and **John is committed to getting the president re-elected**.

characterize commitments in terms of the future discourse actions the commitment excludes. For example, if an agent commits to a belief that p , this precludes future discourse moves that express a belief that is incompatible with p .

This was adequate for the purposes of Hamblin and Gunlogson, but it is too limited in the present context for two reasons. Firstly, the kinds of commitments created by linguistic utterances often concern actions that take place after the current discourse has ended. Secondly, many of the commitments we are interested in here concern *non-linguistic* actions. The commitments resulting from promises and orders are particularly obvious examples: These can of course concern linguistic actions (**Don't say another word!, Say it ain't so!**), but mostly, they do not (**Be at the airport at noon!**).

The discourse commitments of Hamblin and Gunlogson come out as a special case on the perspective taken here: Commitments constrain actions in general, including linguistic actions. To the extent that it is necessary to also represent commitments that are in effect only for the present discourse, we can allow commitments that are only temporary, i.e., that constrain actions only up to a given point in time, such as the end of the current discourse, or the end of the current discourse segment.¹¹

5.3.1 Commitments exclude possible future states of the world

Commitments will be construed as excluding possible future states of the world in our branching-time model. This is somewhat parallel to the way discourse commitments are modeled in Hamblin (1971) and Gunlogson (2008), where commitments constrain the set of 'legal' (Hamblin) or 'expected' (Gunlogson) future states of the discourse. That is, discourse commitments are modeled by excluding all future discourse states at which the commitment is not honored.

This will not do if we model commitments as excluding possible future states

¹¹The latter is what happens when a speaker agrees to a hypothesis 'for the sake of the argument': For the current discourse segment, he commits to act as though he believed the hypothesis, but this commitment is not as enduring and open-ended as the commitment that he would have taken on otherwise.

of the world: Of course, taking on a commitment does not make it impossible to violate the commitment. But there are future states that are indeed made completely impossible by taking on a commitment: Those where the commitment is not honored, and yet, the agent does not count as having violated a commitment.

One-off commitment to action: An illustration

To bring out the intuition, let us consider a somewhat artificial case of a commitment to a one-off action. At time t , John takes on the commitment to raise his hand at a future time t' . After t , there are two kinds of possible future worlds:

1. future worlds in which John raises his hand at t' .
2. future worlds in which John does not raise his hand at t' , and counts as having violated his commitment.

Before t , there was a third kind of possible future, namely those in which John does not raise his hand, yet is not at fault for having violated the commitment to do so.

A necessary complication: Commitments can be rescinded, or canceled.¹² So there are really three kinds of possible futures after t : The two characterized above, plus:

3. future worlds in which the commitment gets rescinded before t' , and John does not raise his hand at t' .

The taking on of the commitment thus can be construed as making impossible all future worlds of the kind characterized below:

4. future worlds in which the commitment does not get rescinded between t and t' in which John does not raise his hand at t' and does not count as having violated the commitment to do so.

¹²In the real world, commitments vary with respect to which agent 'controls' a commitment, i.e., which agent has the right to rescind the commitment: Promises can primarily be rescinded by the promisee, while declarative commitments to belief can often be freely rescinded by the agent who made the commitment in the first place (**I take that back**). I ignore this complication here.

Commitments to courses of actions

Agents do not always commit to single actions, as above, but frequently commit to complex future courses of action. Suppose, for example, that John commits to never drink alcohol again. There is no single future time at which this commitment will be honored or violated. Instead, this commitment will be honored or violated every time John has the opportunity to take a drink.

This complicates matters, as ‘being in accordance with the commitment’ is no longer a stable property. In the case of a one-off action, to be performed at time t' , there was a single decisive point (viz., t') at which the commitment was honored or violated. If John performed the necessary action at t' , at all future times, he was in accordance with his commitment. For open-ended commitments, this is no longer true: If John has not taken a drink up to time t'' , he is in accordance with his commitment, but there are still many possible futures after t'' at which he no longer is, namely, all those at which he takes a drink after t'' .

The notion of ‘having violated the commitment’, however, is still a temporally stable one: If, at any given time, John takes a drink, he will henceforth count as having violated his commitment. For this reason, it is expedient to characterize commitments in general using the notion of *failing to act in accordance with a commitment*, rather than the—*prima facie*—simpler notion of *acting in accordance with a commitment*.

So, informally, for a given property P of actions (such as ‘the action is not a drinking of alcohol’), we explicate ‘commitment to P ’ as follows:

- (5.31) When an agent i takes on a commitment to P at time t , he excludes all possible future worlds w' from becoming actual which are such that, for some t' after t :
- a. in w' , i does not act in accordance to P at t' AND
 - b. in w' , P has not been rescinded prior to t' AND
 - c. in w' , i is not at fault at t' .

This characterization employs the primitive predicate ‘being at fault’, which I will

leave unanalyzed here. Seeing as the commitments in questions are taken to arise on the basis of social conventions, it might be more appropriate to circumscribe it as ‘is considered at fault by the community’, but I shall use the simpler formulation here.¹³

5.3.2 Commitments to beliefs and preferences

The preliminary characterization of the notion of commitment in the previous section was done in terms of predicates of courses of action. In the present section, I will implement this indirectly in the system of dynamic pragmatics, modeling *commitment to belief* (doxastic commitment) and *commitment to preferences* (preferential commitment). This indirection is required because we want commitments to *interact*: Whether or not any given action choice of an agent violates his commitments will, in general, not depend on single commitments of the agent, but on their sum.

Suppose, for example, that an agent is committed to a preference for being at the airport at noon, and also to a belief that, in order to be at the airport at noon, he has to leave home before 10:30. Suppose further that the agent is still at home at 10:45. We want to be able to say that he is in violation of his commitments, but his being at home is compatible with both his preferential commitment and his doxastic commitment, taken individually. It is only *together* that the two commitments are incompatible with his being at home.

I hence introduce predicates *pb* (for ‘public belief’) and *pep* (for ‘public effective preference’), which are indexed to individuals and times, and take propositional arguments. Model-theoretically, these are interpreted using two set-valued functions that specify the beliefs and preferences an agent is committed to, *PB* and *PEP*. I require these to have the following (joint) closure properties.

(5.32) For any i, t, w

- a. $PB_i(t, w)$ is closed under logical inference with an SD45-logic for $pb_{a,t}$.

¹³Ultimately, we probably want to index ‘being at fault’ with the commitment(s) the agent has violated (or rather—to avoid circularity—the commitment-creating events with respect to which he is ‘at fault’). This, again, is a complication I avoid here.

- b. If $p \in \text{PEP}_i(t, w)$ then $\text{pep}_{i,t}(p) \in \text{PB}_i(t, w)$.
- c. If $\text{pep}_{i,t}(p) \in \text{PB}_i(t, w)$ then $p \in \text{PEP}_i(t, w)$.
- d. If $p \in \text{PEP}_i(t, w)$, then $\neg p \notin \text{PB}_i(t, w)$

We use these two sets to interpret the operators, in analogy to the interpretation of B and ep :¹⁴

- (5.33) a. $w \vDash \text{pb}_{i,t}(\phi)$ iff $\phi \in \text{PB}_i(t, w)$.
 b. $w \vDash \text{pep}_{i,t}(\phi)$ iff $\text{PEP}_i(t, w) \vdash_{i,t,w} \phi$

That is, we treat pep as an effective preference structure that only contains maximal elements.

(5.33a) simply ensures that pb behaves like a doxastic operator—in particular, it requires positive and negative introspection. This ensures that:

- (5.34) a. $\text{pb}_{i,t}(\text{pb}_{i,t}(p)) \Rightarrow \text{pb}_{i,t}(p)$
 b. $\text{pb}_{i,t}(p) \Rightarrow \text{pb}_{i,t}(\text{pb}_{i,t}(p))$

(5.33b) says that if an agent is committed to believe that he is committed to prefer p , he is also committed to prefer p —Condoravdi and Lauer (2011)'s **Doxastic reduction for preferential commitment**. (5.34c) is the inverse. While it will not play an essential role in what is to follow, it seems so plausible that I impose it nonetheless.

I model the commitments here using sets of formulas because it allows an easy way to state the closure properties syntactically. The desired logical properties for pb alone could be implemented easily in the standard way, by adding a time-indexed accessibility relation $\mathbb{C}_a(t)$ for each agent a , which is required to be transitive, Euclidean and serial, but ensuring introspection pep and pb would require some additional structure I wish to avoid here.

In the present context, we ignore the possibility of rescission of commitments and require that doxastic and preferential commitments are increasing:

¹⁴I take it to be obvious how the definition of $\sim_{i,t,w}$ has to be adjusted for pep —by replacing reference to the agent's *actual* belief state with reference to his *public* one.

(5.35) If $t < t'$ and $PB_i(t, w) \subseteq PB_i(t', w)$.

(5.36) If $t < t'$ then $PEP_i(t, w) \subseteq PEP_i(t', w)$.

Now we connect these two constructs to at-fault-ness. To this end, we introduce another time-indexed predicate *AtFault* into the pragmatic language. *AtFault* applies to an individual if he counts as having violated a commitment. At-fault-ness is connected to PB and PEP by the following constraint on the interpretation of the *AtFault*-predicate:¹⁵

- (5.37) A model for \mathcal{P}_{Sen} is admissible only if, for all w, t, i : $\langle i \rangle \in I(w)(t)(\text{AtFault})$ iff there is $t' \leq t$:
- (i) $Ag_t(w, t) = i$
 - (ii) $Hap_w(t, t + 1) = a$
 - (ii) $a \notin \text{Opt}(\bigcap PB_i(t', w), PEP_i(t', w), \text{Act}(w, t))$

This just says that an agent is *AtFault* at t if, at some previous time t' , he performed an action that was not consistent with his public beliefs and public preferences at t' .

The DECLARATIVE CONVENTION can now be implemented in the obvious way: Declaratives add their contents to the PB-set of the speaker. However, I do not take PB to *only* contain elements that have been added by explicit utterance events—that would leave the system far too weak. Minimally, it has to contain certain ‘obvious facts’, for example about events that have happened in the course of the conversation, as well as certain basic background assumptions that could not possibly be doubted by a rational agent. In this way, PB is similar to Stalnakerian speaker presuppositions—things that are taken ‘for granted’ automatically are PBs. In particular, I take *utterance events* to be recorded in the PBs of all utterance participants, in the way we have ensured that speakers have perfect knowledge about utterance events in Chapter 2 by means of the PAL-constraint.

¹⁵For this version of the constraint to deal with *failures* to act, we have to construe inaction as the performance of a ‘null action’.

5.4 Modeling conventions of use

We are now almost in the position to implement the normative conventions of use governing the effects of clause types. The only thing that is left to introduce is the modal operator **Result** indexed to events, with the intended interpretation ‘ ϕ is true as a result of e' ’.

(5.38) If e is an event-constant or variable and ϕ a proposition, then $\text{Result}_e(\phi)$ is a formula.¹⁶

Result is interpreted via the function $\text{Res} : W \times T \times T \mapsto \wp(W)$. This function should obey a number of constraints which I will not go into here, save for requiring that it is closed under entailment, and that it is factive:¹⁷

(5.39) If $X \in \text{Res}(w, t, t')$, then $w \in X$.

Now we can capture normative conventions of use for sentence types by imposing additional constraints on admissible models, such as the following:¹⁸

(5.40) **DECLARATIVE CONVENTION constraint**
A model for \mathcal{P}_{Sen} is admissible only if for all φ, e, i, i' : If $\varphi \in \mathcal{L}_{\downarrow}$ and $\text{Hap}_w(I(e)) = \text{utter}(i, i', \varphi)$ then

$$w \models \text{Result}_e(\text{pb}_{i,t}(\varphi))$$

Anticipating the treatment of imperatives in Chapter 6, we can also state the

¹⁶Recall that event constants and variables refer to event *tokens*, which in the current system are simply identified with pairs of times.

¹⁷Another plausible requirement is that **Result** is causally closed in the following sense: If e is causally sufficient for e' and $\text{Result}_{e'}(\phi)$, then $\text{Result}_e(\phi)$. I don't introduce this as an official requirement because modeling causal relationships would require a further complication of the current setup that do not concern us here.

¹⁸Stating these conventions as global constraints on models has the effect that all agents always believe the convention to be in place (and are, in fact, also committed to believe they are). This makes sense, given that we assume knowledge of these conventions to be part of the basic linguistic competence of speakers of a language.

(5.41) IMPERATIVE CONVENTION constraint

A model for \mathcal{P}_{Sen} is admissible only if for all φ, e, i, i' : If $\varphi \in \mathcal{L}_I$ and $\text{Hap}_w(I(e)) = \text{utter}(i, i', \varphi)$ then

$$w \models \text{Result}_e(\text{pep}_{i,t}(\varphi))$$

These constraints encode the conventions in a rather straightforward way: If a speaker utters a declarative or an imperative, the content of his utterance gets added to his doxastic commitment or his preferential commitments, respectively.

We immediately obtain two results: Firstly, if a speaker utters incompatible declaratives, then any future action will make him AtFault, as Opt is undefined if its belief state argument is empty. Secondly, after a speaker utters two imperatives that are incompatible, given his doxastic commitments, any future action will make him AtFault, given the consistency requirement Opt imposes on its preference state argument.

This is of course far too strong in general, but the result relies crucially on the assumptions stated in (5.35) and (5.36), i.e., on the assumption that commitments are non-decreasing. Thus we already capture the fact that if an agent commits to incompatible beliefs or preferences, this requires rescission of a prior commitment.

5.5 Communicating contents, again

We can now show how it comes about that an addressee comes to believe the content of declaratives that a speaker utters. Recall that, in Chapter 2, we already derived this, relying on the following two contextual beliefs of the addressee:

(5.42) *Contextual assumptions: Trusting addressee* =(2.34)

a. 'Honest speaker'

$$B_{Ad,t} \models \text{utter}_i(Sp, \vdash \varphi) \Rightarrow \Box_{Sp,t} \varphi$$

b. 'Informed speaker'

$$B_{Ad,t} \models \Box_{Sp,t} \varphi \Rightarrow \varphi$$

We now replace (5.42a) with the following condition (as with (5.42), this can be seen as a specific instance of a more general belief about the speaker's preferences):

$$(5.43) \quad \text{'Cautious speaker'}$$

$$B_{Ad,t} \models \text{ep}_{Sp,t}(\neg \Box_{Sp,t+1} \varphi \Rightarrow \neg \text{pb}_{Sp,t+1}(\varphi))$$

For simplicity, assume that the speaker chooses only between two actions, and that the addressee knows this:

$$(5.44) \quad \forall w \in B_{Ad,t} : \text{Agt}(w, t) = Sp \text{ and } \text{Act}(w, t) = \{\text{utter}(Sp, Ad, \varphi), \perp\}$$

\perp stands for the 'null action' of doing nothing. That is, we assume that the speaker only faces the choice whether to assert φ or not.¹⁹ Given the declarative convention, we also have, for all worlds v (and hence in particular for all worlds $v \in B_{Ad,t}$):

$$(5.45) \quad B_{Sp,t,v}[\text{utter}(Sp, Ad, \varphi)] \models \text{pb}_{Sp,t+1} \varphi$$

Similarly (assuming Sp is not already committed to φ):

$$(5.46) \quad B_{Sp,t,v}[\perp] \not\models \text{pb}_{Sp,t+1}$$

Assuming that the addressee is uncertain whether the speaker believes φ to be true, there are two kinds of worlds in $B_{Ad,t}$:

- $$(5.47) \quad \begin{array}{l} \text{a.} \quad \text{Worlds } v_\varphi \text{ such that } v_\varphi \models \Box_{Sp,t} \varphi. \\ \text{b.} \quad \text{Worlds } v_{\neg\varphi} \text{ such that } v_{\neg\varphi} \models \neg \Box_{Sp,t} \varphi \end{array}$$

Given (5.45), for v_φ -worlds, we have (5.48a) and (5.48b), which (in the absence of additional relevant preferences) ensures (5.48c), where $\text{Opt}(w)$ is an abbreviation for the outcome of Opt , given Sp 's beliefs and preferences at w .²⁰

¹⁹Of course, in realistic scenarios, there will often be many alternative utterances the speaker could make, and even if we assume that φ is very topical, before the utterance, the addressee likely takes it to be possible that the speaker utters $\neg\varphi$ instead. But for the sake of illustration, it is useful to use a somewhat artificial example.

²⁰That is, $\text{Opt}(w) = \text{Opt}(B_{Sp,t,w}, \text{EP}_{Sp}(w, t), \text{Act}(w, t))$.

- (5.48) a. $B_{Sp,t,v_\varphi}[\text{utter}(Sp, Ad, \varphi)] \models \Box_{Sp,t}\varphi \wedge \text{pb}_{Sp,t+1}\varphi$
 b. $B_{Sp,t,v_\varphi}[\perp] \models \Box_{Sp,t}\varphi \wedge \neg \text{pb}_{Sp,t+1}\varphi$
 c. $\text{utter}(Sp, Ad, \varphi) \in \text{Opt}(v_\varphi)$

For $v_{\neg\varphi}$ -worlds, we have instead:

- (5.49) a. $B_{Sp,t,v_{\neg\varphi}}[\text{utter}(Sp, Ad, \varphi)] \models \neg\Box_{Sp,t}\varphi \wedge \text{pb}_{Sp,t+1}\varphi$
 b. $B_{Sp,t,v_{\neg\varphi}}[\perp] \models \neg\Box_{Sp,t}\varphi \wedge \neg \text{pb}_{Sp,t+1}\varphi$
 c. $\text{Opt}(v_{\neg\varphi}) = \{\perp\}$

But, given the **Belief in optimal action choice** constraint, this ensures that only v_φ -style worlds are such that $\text{Hap}(t, t+1) = \text{utter}(Sp, Ad, \varphi)$, and so:

- (5.50) $B_{Ad,t}[\text{utter}(Sp, Ad, \varphi)] \models \Box_{Sp,t}\varphi$

And, assuming ‘Informed speaker’:

- (5.51) $B_{Ad,t}[\text{utter}(Sp, Ad, \varphi)] \models \varphi$

5.6 Conclusion

This chapter has introduced a good deal of machinery, which was illustrated with a rather simple example of a pragmatic inference: The inference to the truth of what the speaker said. The usefulness of the machinery will reveal itself in the chapters to follow.

It is worth reflecting, however, on what we have done. We started in Chapter 2 with a system that let us show that an addressee who makes the ‘Honest speaker’ assumption will learn that the speaker believes in the truth of what he says. In Chapter 3, I demanded justification for this contextual belief, on the basis that it is warranted only for declaratives, but not for other sentence types.

Now, at the end of the present chapter, we have again derived the same result, replacing the contextual assumption ‘Honest speaker’ with the contextual assumption ‘Cautious speaker’. A critical reader may wonder—what have we gained? We

have replaced one assumption with another. Where is the justification for ‘Cautious speaker’?

Let us compare the two contextual assumptions, repeated below.

(5.52) ‘Honest speaker’

$$B_{Ad,t} \models utter_t(Sp, \vdash \varphi) \Rightarrow \Box_{Sp,t} \varphi$$

‘A speaker will utter a declarative with content φ only if he believes it to be true.’

(5.53) ‘Cautious speaker’

$$B_{Ad,t} \models ep_{Sp,t}(\neg \Box_{Sp,t+1} \varphi \Rightarrow \neg pb_{Sp,t+1}(\varphi))$$

‘A speaker prefers not to commit himself to the proposition φ unless he believes it to be true.’

‘Honest speaker’ makes essential reference to the *clause type* of the uttered sentence: It is an assumption about when sentences in \mathcal{L}_+ will be uttered. ‘Cautious speaker’ makes no reference to sentences of the object language *Sen* at all, but only talks about propositions (which it happens to identify by sentences of *Prop*).

Justification of ‘Honest speaker’ hence requires appeal to narrowly linguistic concepts, while justification of ‘Cautious speaker’ does not. It is because of interlocutors’ linguistic knowledge that ‘Honest speaker’ is justified.

At the end of Chapter 4, I have proposed that the linguistic knowledge that justifies ‘Honest speaker’ (in appropriate contexts) is related to the *dynamic effect* that is conventionally associated with declarative sentences, and that this effect consists in the creation of *commitments* to choose one’s actions in a particular way. To make this hypothesis precise, the current chapter has introduced a model of action choice (making as few assumptions as necessary about how it works, by specifying general constraints on *Opt* relative to the chosen representation of preferences and beliefs), and then gone one to specify a model of the notion of commitment suitable for the statement of clause-typing conventions.

The usefulness of the apparatus introduced in this chapter is, of course, not exhausted by explaining the simple pragmatic inference derived in the last section.

This is just the beginning. We now have a fairly general framework at our disposal that can be applied to a variety of pragmatic phenomena. In Chapter 6, I apply the framework to the other two main clause types, imperatives and interrogatives. Chapter 7 shows how the current set-up can account for the fact that explicitly performative utterances have the result of making the content of the uttered sentence true. Chapter 8 discusses exclamatives, arguing that they should not be viewed as giving rise to commitments. And finally, Chapter 9 applies the framework to model conversational implicatures.

Chapter 6

Commitments to preferences

The previous chapters were largely concerned with declarative sentences and their effects. Other sentence types have been mentioned, but mainly in order to provide a contrast to declaratives, which were argued to be associated with a convention that ensures that a speaker who utters a declarative comes to be committed to believe that that content of the declarative is true.

The present chapter is concerned with sentences that commit their speakers to a *preference*. The main examples are *imperatives*. According to the analysis adopted in this chapter, an utterance of an imperative conventionally commits the speaker to a preference for the imperative to become fulfilled (Section 6.1). Section 6.2 discusses a class of declarative sentences that have imperative-like uses. While utterances of these sentences, like all declarative utterances, conventionally give rise to a commitment to a belief, I will argue that this doxastic commitment can be viewed as giving rise to a preferential commitment indirectly. Section 6.3 sketches how the emerging system of clause-typing could be extended to the third ‘major’ sentence type: interrogatives. Interrogatives are analyzed on the model of imperatives, as being conventionally associated with a preferential commitment of a very particular kind. The chapter concludes with Section 6.4, which discusses the question what the denotation of sentences of various types should be and revisits the question whether information relating to (sentential) force should be represented in the compositionally-determined denotation of sentences.

6.1 Imperatives

A central problem for an analysis of imperatives is the PROBLEM OF FUNCTIONAL HETEROGENEITY, which was first appreciated in its entirety by Schmerling (1982): Imperatives come with a wide variety of uses, and it is difficult, at least at first glance, to find a unifying feature of all of them. Consider the examples in (6.1), repeated from page 46.

- | | | | |
|-------|----|---|---------------|
| (6.1) | a. | Stand at attention! | (COMMAND) |
| | b. | Don't touch the hot plate! | (WARNING) |
| | c. | Hand me the salt, please. | (REQUEST) |
| | d. | Do the right thing! | (EXHORTATION) |
| | e. | Take these pills for a week. | (ADVICE) |
| | f. | Please, lend me the money! | (PLEA) |
| | g. | Get well soon! | (WELL-WISH) |
| | h. | Drop dead! | (CURSE) |
| | i. | Okay, go out and play. | (PERMISSION) |
| | j. | Okay then, sue me, since it's come to that. | (CONCESSION) |
| | k. | Have a cookie(, if you like). | (OFFER) |

Commands such as (6.1a) are the stereotypical (though not necessarily typical) use of imperatives. They have two features, which they share with other 'directive' uses, such as requests (6.1c), warnings (6.1b) and exhortations (6.1d): (i) They act as an *inducement* for the addressee to fulfill the imperatives and (ii) they express that the speaker *desires*, in some sense, the addressee to fulfill the imperative.

However, neither of these properties is present on all uses: A well-wish like (6.1g) and an ill-wish like (6.1h) do not act as inducements, they merely express a desire. And permission and invitation uses like (6.1j-k) do not appear to express a speaker desire at all, though they may act as inducements in some sense. This is even more obvious in cases of disinterested advice, as in (6.2).

(6.2) [Strangers in the street of Palo Alto]

A: Excuse me, how do I get to San Francisco?

B: Take the north-bound Caltrain.

B's utterance does not, in any obvious sense, express a desire on *B*'s part, even though, given *A*'s mutually-known, salient goal, it will likely act as an inducement for *A* to get on the Caltrain.

The problem of functional heterogeneity makes it particularly difficult to assign imperatives a uniform sentential force—i.e., a uniform, conventionally-determined essential effect that is present for all utterances of imperatives, as it is not obvious what this uniform effect should be.

Condoravdi and Lauer (2012) argue, however, that we *can* specify such a uniform effect, and derive the varied uses of imperatives in context. The following sections summarize the proposal, and show how the system of dynamic pragmatics can be used to demonstrate, on a very fine-grained level, how specific uses of imperatives arise through the interplay of the conventionally specified effect of imperatives and features of the context in pragmatic, interactional reasoning.

I will not follow the strategy of Chapter 4 and try to discuss various alternative ways the force of imperatives could be specified. The main reason is that an excellent exhaustive review of the literature exists, viz., that in Kaufmann (2012). I would have little to add to Kaufmann's discussion of previous proposals. As Kaufmann's discussion shows, few of the proposals predating her book are even initially plausible once one appreciates the functional heterogeneity of imperatives. That leaves the two main current alternative approaches, Kaufmann's own and that of Portner (2005, 2007). I refer the reader to Condoravdi and Lauer (2012) for a discussion of these, as well as an in-depth comparison with the analysis proposed here. A noteworthy feature of both approaches is that they take imperatives to be *semantically underspecified* much in the way that modals like **must** are in analyses following Kratzer (1981), and they explain functional heterogeneity as arising from this underspecification. The analysis proposed here and in Condoravdi and Lauer (2012), by contrast, assumes no semantic underspecification.

6.1.1 Imperatives as creating commitments to preferences

Condoravdi and Lauer (2012) propose that the essential effect of imperatives is the creation of a commitment to a preference:

(6.3) IMPERATIVE CONVENTION (NL statement)

When a speaker utters an imperative that has the content φ , he thereby commits himself to prefer φ to be actualized.

This assumes that the content of imperatives is a proposition—the proposition that is true if the imperative is fulfilled, i.e., that (6.4a) roughly has the same denotation as (6.4b):

- (6.4) a. Leave!
b. You will leave.

This assumption is not at all essential. The analysis of imperatives proposed here is compatible with a wide range of possible denotations for imperative, as I show in Section 6.4. In the meantime, I will assume such propositional denotations for the sake of concreteness. That is, I assume that

$$(6.5) \quad \llbracket !\phi \rrbracket^{Sen} = \llbracket \phi \rrbracket^{Prop}$$

With this, we can state (6.3) as a constraint on pragmatic models (this definition was anticipated already in Section 5.3.2 above):

(6.6) IMPERATIVE CONVENTION constraint (5.41)
A model for \mathcal{P}_{Sen} is admissible only if for all φ, e, i, i' : If $\varphi \in \mathcal{L}_!$ and $\text{Hap}_w(I(e)) = \text{utter}(i, i', \varphi)$ then

$$w \models \text{Result}_e(\text{pep}_{i,t}(\varphi))$$

In view of the stereotypical uses of imperatives, i.e., commands, this proposal may seem surprising: Imperatives commit the *speaker* to a preference? Isn't the main

point of a command-imperative like (6.7) to get the *addressee* committed to perform the commanded action?

(6.7) Be at the airport at noon!

Indeed, on command uses, this is generally the *point* of uttering an imperative, but we deny that it is the conventionally-determined *effect* of imperative utterances.

One reason for this is conceptual. It would be quite strange if an utterance, in virtue of *linguistic* convention, could commit the addressee to something. At best, it could *attempt* to commit the addressee to something. The other reason is empirical and lies in the problem of functional heterogeneity: There are many uses, such as advice, permission, and wish uses, for which it is implausible to say that they involve an addressee commitment, even an attempted one.

Assuming that imperatives give rise to speaker-commitments to preferences brings into sharp focus, however, the flip-side of the problem of functional heterogeneity. While imperatives can have a wide range of uses, there are others that are categorically excluded. Most notably, imperatives cannot be used to promise or threaten:

(6.8) a. Get a promotion!

(cannot be a promise that the speaker will ensure that the addressee gets a promotion)

b. Lose your job!

(cannot be a threat that the speaker will ensure that the addressee loses his job)

An analysis that hypothesizes that imperatives commit the speaker to a preference for the imperative to be fulfilled highlights this problem, as promises arguably *are* just utterances that commit the speaker to a preference for what is promised (cf. Chapter 7). As Condoravdi and Lauer (2012) argue, however, the main alternative accounts of imperatives, those of Kaufmann (2012) and Portner (2005, 2007), face the same problem, at least to a certain degree. The commitment-based analysis

merely makes the issue particularly obvious.

I shall not dwell on this problem here, but refer the reader to Condoravdi and Lauer (2012), where we propose to solve it by assuming that, in addition to the creation of a preferential commitment, imperatives come with a secondary effect which limits the involvement of the speaker in bringing about the truth of the proposition denoted by the imperative, ruling out promise and threat uses.

In the following, I briefly summarize some desirable consequences of the proposed analysis. In Condoravdi and Lauer (2012), we argue that all these consequences are *necessary* for a satisfactory analysis of imperatives, and also compare the commitment-based analysis with that of Kaufmann and Portner, showing that neither meets all the desiderata that we argue for.

Contextual inconsistency As pointed out in Section 5.3.2, a consequence of the current analysis is that subsequent imperatives by the same speaker are required, just like declaratives, to be consistent. This is a welcome prediction.

Here is why: Imperatives require retraction or revision even when they are not used with the same ‘force’. This is observed by Portner (2007, p. 367): “it’s odd to give conflicting imperatives even when they are of different subtypes (unless you have changed your mind, of course), as shown in example [(6.9)]. This pair of sentences cannot be coherently uttered by a single speaker.”

- (6.9) Stay inside all day! (order)
 #Since you enjoy the nice weather, go out and play a little bit. (suggestion)

This does not follow from the fact that the two imperatives, intuitively, convey that the speaker wants them to be fulfilled. As I noted before, it is quite consistent to have inconsistent desires—unless one decides to act on them:

- (6.10) I want to stay inside to avoid getting a sunburn, but I also want to go out and play!

Why, then, can (6.9) not be taken to be an indication of such inconsistent desires?

On the present account that is because (6.10) involves commitment to *effective* preferences, which need to be consistent.¹

Speaker endorsement A speaker who utters an imperative with content p cannot, without being subject to blame, act so as to prevent the realization of p . This is something that is universally true for all uses of imperatives, even ones that intuitively do not involve a desire of the hearer or deontic authority, such as the advice-use in (6.2). As a consequence, certain continuations are infelicitous after an imperative, regardless of which ‘force’ the imperative has:

- (6.11) a. Go, # but I won’t allow you to.
 b. Go, # but I don’t want you to.

Declarations that the speaker would prefer it if his imperative were not fulfilled are possible, but only if they explicitly concern non-effective preferences, as is arguably the case for **wish**:

- (6.12) Okay, go to the party, but I wish you would not.

Functional heterogeneity The main idea underlying a uniform analysis of imperatives (instead of one relying on underspecification, such as those of Kaufmann (2012) and Portner (2007)) is that the distinctive features of the varied uses of imperatives can be explained by interaction of the uniform dynamic effect of imperatives together with contextual conditions and pragmatic reasoning. The rest of this section illustrates how we can model this interaction in the system of dynamic pragmatics.

The presentation will focus on two exemplary classes of uses. On the one hand, Section 6.1.2 discusses uses that are *directive* in a narrow sense, such as

¹Kaufmann (2012)’s account arguably is unable to derive the consistency requirement. Portner (2007) appears to intend to simply stipulate it as a requirement on ToDo-lists—but in Portner (2012), the author suggests that allowing for inconsistency is the right way to deal with PERMISSION uses. If so, then Portner’s account cannot jointly account for PERMISSION uses and the consistency constraint. See Condoravdi and Lauer (2012) for discussion.

orders, requests, and pleas. These are the stereotypical uses of imperatives, and it is not immediately obvious that our analysis predicts them. On the other hand, Section 6.1.3 focuses on *advice* uses. These are particularly interesting from the perspective of dynamic pragmatics, as they involve quite intricate inferencing of the Gricean sort, whereas directive uses do not. Relatedly, advice uses are somewhat unique among the uses of imperatives in that their main point often seems to be to *convey information* (about how a certain goal can be reached), which is the stereotypical purpose of *declaratives*.

6.1.2 Deriving directive uses

(6.13) illustrates some directive uses of imperatives. Each example is paired with a reportatively used declarative that can be used to describe the corresponding imperative after the fact.

- (6.13) a. (i) Stand at attention! (COMMAND)
 (ii) He ordered me to stand at attention.
 b. (i) Hand me the salt, please. (REQUEST)
 (ii) He requested to be passed the salt.

We want to answer two questions about directive uses: (i) Under which circumstances can an utterance of an imperative be described with a directive verb of saying such as **order** and **request**? (ii) Why do utterances of imperatives, in these contexts, act as an inducement for the addressee to fulfill them? (Why can they create an effective preference of the addressee for fulfilling them?)

Imperatives and directive verbs of saying

I take **order** and **request**² to have the same truth-conditional content, summarized in (6.14), and assume they differ only in their presuppositions. Roughly, **order**

²In Condoravdi and Lauer (2012), we also took **warn** to be equivalent to these directive predicates in terms of its truth-conditional content, but I am no longer sure that this is adequate—in particular, I think that the preference expressed by a **warning** does not necessarily require that the addressee form an effective preference. Warnings seem more akin to cases of **advice** in this way.

presupposes that the speaker believes himself to have deontic authority over the addressee, whereas **request** presupposes that the speaker takes it to be likely that p does not interfere with the current effective preferences of the addressee.

$$(6.14) \quad \text{utter}_u(a) \text{ and } \text{Result}_u(\text{pep}_a(\text{ep}_b(p)))$$

The assumed dynamic effect of imperatives means (6.15) will be true after the utterances of imperative with content p :

$$(6.15) \quad \text{utter}_u(a) \text{ and } \text{Result}_u(\text{pep}_a(p))$$

Our task is now to specify properties of the context under which (6.15) entails (6.14). The condition I want to propose here is the following: Directive uses arise only if the addressee is taken to have *control* about p , that is, if it is taken for granted that p will be realized if and only if b 's preferences are such that he would choose it.

$$(6.16) \quad \text{PB}_a(w, t) \models p \Leftrightarrow \text{ep}_b(p)$$

(6.16) is indeed sufficient to ensure that (6.15) entails (6.14): If it is true, then p and $\text{ep}_b(p)$ are indistinguishable in the speaker's public belief state, and hence for any preference structure PEP in which p is a maximal element: $\text{PEP} \sim_{a,t,w} \text{PEP} + \text{ep}_b(p)$.

So if (6.16) holds, then an utterance of an imperative automatically ensures that the truth conditions of directive predicates are satisfied. But it is a rather strong condition. Is it justified to assume it holds whenever an imperative is used with directive force? I will try to motivate the two directions of the implication separately.

Motivating $p \Rightarrow \text{ep}_b(p)$. This direction can be motivated rather easily. It says that p will be actualized only if b 's preferences require it, i.e., that the actualization of p depends on b 's *choice*. This is a natural assumption for many agentive predicates, e.g., the ones exemplified in (6.17).

- (6.17) a. Close the window!
 b. Leave!
 c. Pay your taxes!
 d. Call me!

Not all directive uses of imperatives involve predicates that are necessarily agentive. In particular, such directive uses can happen with *stative* predicates and those that involve refraining from actions (which arguably are stative, too), as in (6.18).

- (6.18) a. Sit still!
 b. Don't say another word!

It seems plausible, however, to say that a speaker will utter such imperatives (as directives) only if he takes it for granted that they will not become realized unless the addressee forms the correct effective preference.³

Motivating $p \Leftarrow ep_b(p)$. It is perhaps not quite as obvious that agentivity, insofar as it is required for directive imperatives, also ensures the opposite implication, i.e., that p will be actualized if b effectively prefers it. This is certainly plausible for imperatives such as **Leave** or **Shut up**, but does it hold in general? There is some reason to think that when an imperative is used directly, it does. Suppose the speaker has authority over the addressee, and wants him to be at the airport at noon, but he knows that whether the addressee can be there in time depends on factors like the traffic conditions, which are not under the addressee's control. In that case, it seems a speaker would use (6.19a) rather than (6.19b).

- (6.19) a. Try to be at the airport at noon!
 b. Be at the airport at noon!

It does seem intuitively correct to say that if a speaker uses (6.19b) (as a command or request), he indeed must presuppose that the addressee can ensure that he is

³Perhaps it would be adequate to add this speaker presumption as another presupposition for **order**, **request**, etc.

at the airport at noon. This strongly suggests that, for directive uses, it is indeed necessary that p is considered, in the given context, to be under the control of the addressee in the strong sense demanded by (6.16). I have to concede, however, that I cannot exhaustively demonstrate at present that imperatives can *only* have directive force if (6.16) holds. Observations like (6.19) suggest that this is so, but it might turn out that while (6.16) is a sufficient condition for directive uses, it is not a necessary one.

I want to stress that in the present account (6.16) is just a *contextual condition* which may or may not be in place in a given context. It is *not* a feature conventionally associated with imperatives. In the past, (6.16) has sometimes been taken to be a general property of all utterances of imperatives, as in Belnap and Perloff (1990)'s *imperative content thesis* (for these authors, the term 'agentive' implies (6.16)):

(6.20) IMPERATIVE CONTENT THESIS (Belnap and Perloff 1990, p. 173)

Regardless of its force, the content of every imperative is agentive.

But this is too strong as a requirement for all imperative uses: Recall that imperatives (at least in English) can be used to express mere desires, as in well-wishes, 'absent wishes':⁴

- (6.21) a. Get well soon!
 b. [On the way to a blind date, to oneself]
 Be blond/rich/nice!

This indicates that it is wrong to assume that the IMPERATIVE CONTENT THESIS captures a conventional fact about imperatives. Rather, it articulates a contextual

⁴The requirement is also too strong in order account for 'advertising imperatives', like (i) and (ii), as Paul Kiparsky (p.c.) points out:

- (i) Win a trip to Finland!
 (ii) Be the first on your block to own a Tesla!

condition that may or may not be in place, and that is necessary for directive uses only. The set-up of the dynamic pragmatics lets us see how contextually-assumed agentivity results in directive uses: Directive predicates like **order** and **request** apply to imperative utterances only if the speaker of the imperative takes it for granted that (6.16) holds.

Directive imperatives act as inducements to action

Why do directive imperatives (i.e., imperatives with contents that require an addressee choice for their realization) typically act as *incentives to action*? This will depend on socio-cultural circumstance: An imperative functions as an **order** only if the speaker has deontic authority over the addressee, which we can now cash out as follows:

- (6.22) An agent a has deontic authority over another agent b , with respect to p if $\text{pep}_{a,t}(p)$ implies that, at t , b is (socially/legally/institutionally/. . .) obligated to act though he prefers p .⁵

If the speaker does not have deontic authority in this sense, the extent to which the addressee will be induced to fulfill the imperative will depend in complex ways on his other preferences. In many cases of requests, they will be mediated by a desire to help the speaker to fulfill (publicized) goals, which may be motivated in various ways—by a desire to stay in the good graces of the speaker, for example, or by genuine concern for a friend’s well-being.

In the limiting case, though, why would a request made by one stranger to another ever induce the addressee to act? I propose, again following Condoravdi and Lauer (2012), that agents are generally *cooperative-by-default*, in the sense defined in (6.23).

⁵I do not treat permission and concession uses here, but this dependence of the obligation on a public preference will play an instrumental role in accounting for those in the current set-up—permission for p can be viewed as the retraction or modification of the permitter’s existing preferential commitment for $\neg p$. Given the characterization of deontic authority employed here, such a rescission or modification will remove the prohibition against p .

(6.23) COOPERATIVITY BY DEFAULT

An agent *a* is cooperative-by-default iff he adds any topical goal *g* of another agent to his effective preference structure.

(6.23) is very weak: It only requires that topical goals of other agents be added to the effective preference structure, but not that they be ranked in a particular way. The effect this has is that an agent will not thwart another agent's goals without reason, i.e., randomly. The reason may be entirely selfish—laziness, say—but (6.23) requires that, if an agent does not act to help others fulfill their publicized goals, he must have *some* selfish preference motivating this. To realize how weak this principle is, observe that the reason may even be pure antagonism: An agent who fails to aid another simply because he takes pleasure in seeing others' goals thwarted can still be cooperative-by-default, if he ranks his desire to see others suffer over the goals of others.

(6.23) ensures that if the action requested by a speaker does not 'cost' the addressee anything, i.e., if it does not interfere with his private preferences, he will act so as to fulfill the request. If we assume (6.23) is generally true of (social) agents, we can hence explain why a speaker might have reason to assume he can get a complete stranger to move by uttering (6.24).

(6.24) Step out of the way, please.

6.1.3 Deriving advice uses

Advice uses come in two kinds: Those in which it is taken for granted that the speaker and addressee share a salient goal, and those where this assumption is not warranted. (6.25) is an example of the first kind, assuming that a conversation between doctor and patient presupposes that both want the addressee to be healthy.

(6.25) [Doctor to patient]
Take these pills for a week.

(6.26) is an example of the second kind: Between strangers, it is not plausible to assume that they share a topical goal.

- (6.26) [Strangers in the streets of Palo Alto]
 A : Excuse me, how do I get to San Francisco?
 B : Take the north-bound Caltrain.

Here, I want to focus on the first kind of case, supposing that the second kind can be assimilated to the first if we assume agents to be cooperative-by-default.

Advice cases are quite interesting from a pragmatic perspective, because they involve an imperative that does not act as an inducement to action (the way an order, request or plea does), but also *gives information* about the satisfaction of a goal: (6.26a) makes it so that the addressee can conclude something along the lines of (6.27), and that may well be the *main purpose* of the speaker in uttering the imperative.

- (6.27) Taking the pills is the best action to further the goal of curing the addressee's current illness.

We want to understand how this happens. Let *pills* be a proposition letter standing for the proposition that the addressee takes the pills and *cure* be a proposition letter standing for the proposition that the addressee gets cured. Then we can symbolically represent (6.27) as:

- (6.28) *Best(cure, pills)*

(6.28) stands for a possibly quite complex set of facts involving medical knowledge and standards of certainty, etc., so it would be futile to define it as a proposition in our current system. I use it as an abbreviation for the natural language statement in (6.27) here.

The addressee will come to believe (6.27) through an utterance of *!pills* if the following assumptions hold throughout his belief state.

- (6.29) a. SINCERITY
The doctor S will commit to a preference for φ only if he has it.
 $\forall\varphi : \text{pep}_S(\varphi) \Rightarrow \text{ep}_S(\varphi)$
- b. GOAL
The doctor S shares the (topical, salient) addressee's goal of curing his current illness.
 $\text{ep}_S(\text{cure})$
- c. NON-INTERFERENCE
The doctor has no other (topical, salient) preferences relevant to the addressee's taking the pills.
- d. EXPERTISE
If the doctor thinks that φ is the best way to realize *cure*, it is.
 $\forall\varphi : B_S(\text{Best}(\text{cure}, \varphi)) \Rightarrow \text{Best}(\text{cure}, \varphi)$

Assuming that the addressee's pre-utterance information state B_A supports all these conditions, we can derive (6.30d) (omitting all temporal parameters):

- (6.30) a. $B_A[\text{utter}(S, \ulcorner!pills\urcorner)] = \text{Result}_u(\text{pep}_S(\text{pills}))$ (DECL. CONVENTION)
b. $B_A[\text{utter}(S, \ulcorner!pills\urcorner)] = \text{pep}_S(\text{pills})$ (Result is factive)
c. $B_A[\text{utter}(S, \ulcorner!pills\urcorner)] = \text{ep}_S(\text{pills})$ (SINCERITY)
d. $\forall w \in B_A[\text{utter}(S, \ulcorner!pills\urcorner)] :$
 $\text{EP}_S(w) \sim_{S,t,w} \text{EP}_S(w) + \text{pills}$ (def. ep)

Let us make, for a second, the (unrealistic) simplifying assumption that B_A also supports the assumption that the speaker has *no* effective preference other than *cure*. Then (6.30d) comes down to (6.31):

- (6.31) $\forall w \in B_A[\text{utter}(S, \ulcorner!pills\urcorner)] :$
 $\langle\{\text{cure}\}, \emptyset\rangle \sim_{S,t,w} \langle\{\text{cure}, \text{pills}\}, \emptyset\rangle$

(6.31) is true if, according to the speaker's information state $B_S(w)$, there is no decision situation in which jointly satisfying *cure* and *pills* requires a different action than satisfying *cure* alone. Assuming that the speaker believes that he has

the ability to ensure *pills* at all⁶), this means that he must believe that his preference for *cure* is best satisfied by ensuring *pills*, that is, the speaker must believe that $Best(cure, pills)$.

Things would not change if the addressee takes it to be possible that the speaker has other effective preferences, as long as these preferences are orthogonal to taking the pills—i.e., that none of his other preferences plays a role in action choices having to do with ensuring *pills*. Indeed, *A* may realistically assume that there are many such preferences—e.g., the doctor might prefer to have a beer at the end of the day, that his daughter gets into a good college, that England wins the world soccer cup, and so forth. None of these preferences (presumably) are relevant in choosing whether to ensure *pills* becomes true.

So, by NON-INTERFERENCE, we can conclude that

- (6.32) a. $B_A[\text{utter}(S, \ulcorner !pills \urcorner)] \vDash B_S(Best(cure, pills))$
 b. $B_A[\text{utter}(S, \ulcorner !pills \urcorner)] \vDash Best(cure, pills)$ (by EXPERTISE)

But (6.32b) is just what we wanted to derive: If an addressee has an epistemic state B_A which supports our contextual assumptions, then observing an utterance of *!pills* will make him believe that the best way to cure his current illness is to take the pills. The contextual assumptions are rather specific, of course, but we can now investigate what happens if we selectively give them up.

For example, let us assume that the addressee does not trust the medical knowledge of the doctor, i.e., we have all assumptions in place except for EXPERTISE. We no longer predict that (6.32b) holds, but we still predict that (6.32a) holds: The addressee learns that the doctor *believes* that taking the pills is the best way to cure him.

Perhaps more interestingly, let us hypothesize that the addressee fails to believe in NON-INTERFERENCE. Concretely, he knows (or takes it to be possible) that the

⁶That is, that he believes he can convince the addressee to take the pills. Note that we need not assume here that (the addressee believes that) the speaker believes that his current utterance will suffice to convince the addressee. The speaker could believe that, but he could also simply take it to be possible to convince the addressee with a later utterance or other action—e.g., by showing him a medical study that establishes the efficacy of the pills.

doctor gets a kickback from the manufacturer of the pills whenever a patient fulfills a prescription for them. This alone would not change things, if the patient still believes that the doctor's desire for money is kept strictly dominated (in virtue of his professional ethics) by his desire to cure his patient. But let us assume—violating belief in NON-INTERFERENCE—the patient takes it to be possible that the doctor's desire for getting the kickback is active, i.e., that:

$$(6.33) \quad \exists v' \in B_A(w) : v \vDash \text{ep}_S(\textit{kickback})$$

Given that making the patient take the pills will make *kickback* true, an (undominated) effective preference for *kickback* is interfering. In *v*-type worlds, we cannot conclude from (6.34) that the speaker believes that taking the pills is the best way to cure the patient.

$$(6.34) \quad \text{EP}_S(v) \sim_{B_S(v)} \text{EP}_S(v) + \textit{pills}$$

The doctor's decision for *pills* in *v* might also be motivated by his desire for *kickback*. So if the patient is suspicious in this way, we derive the weaker conclusion in (6.35):

$$(6.35) \quad B_A[\text{utter}(S, \ulcorner \textit{pills} \urcorner)] \vDash \textit{Best}(\textit{cure}, \textit{pills}) \vee \text{ep}_S(\textit{kickback})$$

This accords with intuition: If the patient thinks that the doctor has incentive to prescribe him the wrong pills, he will come to believe that either taking the pills will cure him, or the doctor wants to profit from his (presumed) gullibility.

We need not make such sinister assumptions. We derive a similar effect if we assume that the doctor and patient *share* another benign preference that is 'interfering'. Suppose, for example, that the patient does not have health-insurance (or large amounts of spare money), and hence can only afford generics. This is a benign interfering preference, but it is an interfering preference: The patient will no longer conclude (necessarily) that taking the pills is the best way to get him cured, but rather that taking the pills is the best way to get him cured with generics—there might be a medication that is better for him, but does not have a generic equivalent. So we get the strong implication that the patient comes to believe the pills to be

the best way to cure him only if we assume him to believe there are no interfering preferences—otherwise, he will believe that taking the pills is the best way to cure him, taking into account his other preferences. This again accords with intuition.

We thus see how the hearer's inference depends in complex ways on his assumptions about speaker beliefs and preferences (and whether these preferences are well-aligned with his own). We can now generalize the conditions under which we expect an imperative to give advice relative to a salient shared goal:

- (6.36) An addressee of an imperative $p!$ will conclude that p is the best way to satisfy g if he believes the following:
- a. The speaker and the addressee share the goal g .
 - b. The speaker does not have any other effective preference that is plausibly relevant to decisions about p .
 - c. The speaker is an epistemic authority on ways to achieve g .
 - d. The speaker will only commit to preferences he actually has.
 - e. The speaker takes it to be possible that there is a way to convince the addressee to perform p .

Salience, topic-hood and the QUD

The foregoing assumed that there is only one shared goal between the interlocutors, and motivated this by assuming it is the only salient or relevant one. It would be desirable if our model included a mechanism that explains (or at least constrains) how this happens. We only have to enrich the utterance situation slightly to make this assumption worthy of exploration. Doctors and patients may have a number of (shared) preferences in addition to curing the patient's current illness (if it is even established that there is a single illness to be cured). For example, it is reasonable to assume that they also desire to prevent future illnesses that may be unrelated to the current one. Even if such a background goal exists, a patient might well (reasonably!) infer that the doctor's utterance of **Take these pills for a week!** pertains to the current illness—but why?

I believe the correct answer has indeed to do with salience, and perhaps topic-hood of the conversation. The doctor will not utter his imperative out of the blue; it likely is made as part of a longer conversation about how to cure what ails the patient right now. Developing a mechanism to model topic-hood is well beyond the scope of this thesis.

I want to briefly mention a dominant approach to capture something like topic-hood or salience, and explain why I don't think it helps with the problem we face here. In *question under discussion* models (Roberts 1996, Ginzburg 1996), every utterance is taken to address a salient question that is represented as the top-level element of a QUD stack (or a more general structure—Ginzburg allows the order of the possible QUDs to be partial). Perhaps we should integrate such widely-employed representational parameters into our model?

This seems tempting, for isn't the tendency to infer that taking the pills will cure the current illness particularly strong if the doctor utters it in response to an overt question like (6.37)?

(6.37) Doctor, what can I do to fix this?

Indeed, it is. And given that the QUD model allows the question that an utterance addresses to be implicit, we could hypothesize that even if the patient does not utter an interrogative like (6.37), but only describes his problems, a QUD like (6.37) can be inferred.

But it is not clear what we gain from this assumption. Why does (6.37)—whether it is implicit or overt—exclude other possible preferences from consideration? What's worse, we do not even always *want* to exclude other possible preferences—recall the preference for only taking generic medications above. So if the QUD model is to have any bite (in the sense that if the QUD is (6.37), the only preference taken into account is the preference for fixing the current problem), we would need to assume that the QUD in the generic scenario is instead (6.38).

(6.38) Doctor, what can I do to fix this without paying for non-generics?

This may not be implausible as a QUD, but the question remains: How does the QUD model help us to understand which considerations should enter into the interpretation of the imperative utterance and which should not? I submit that it does not, unless it is paired with a substantive theory about how QUDs evolve in dialogue, and how they relate to conversational-external goals of the interlocutors.

This is not to say that the preference-based model laid out in this thesis is *any better* than the QUD model in telling us how salience and topic-hood influence which preferences and beliefs are taken into account at any given point in the conversation—it plainly is not. But it is also not clear how a QUD-less preference-based model is any *worse*. While QUDs intuitively relate to topic-hood (and may be helpful in explaining other information-structural properties), assuming that the QUD is (6.38) is not any more illuminating than hypothesizing that the only two preferences taken into account in the interpretation of the imperative are those for fixing the current problem and only buying generic medications.

So I shall not extend the system of dynamic pragmatics with a QUD representation, and leave the proper way to model salience and topic-hood in the system for future research.

6.1.4 Conclusion

In this section, I have illustrated how the system of dynamic pragmatics helps us to spell out, in a quite detailed fashion, how some of the varied uses of imperatives arise on the basis of a uniform dynamic effect.

A theory such as that of Condoravdi and Lauer (2012) arguably *requires* a rich theory of pragmatic inference to account for the functional heterogeneity of imperatives. A theory that instead relies on semantic underspecification, such as that of Kaufmann (2012), may be seen as less dependent on pragmatic inference, but this comes at a price. In order to correctly account for advice uses, Kaufmann (2012) imposes the following disjunction as one of a number of felicity conditions on utterances of imperatives (c is the context of utterance, g is the ‘ordering source’ parameter of a Kratzerian modal, Δ is a function that specifies the decision problem

(if any) that the interlocutors are trying to resolve in a context):

(6.39) **Ordering source restriction** (Kaufmann 2012, p. 162)

either (i) in c there is a salient decision problem $\Delta(c) \subset \wp(W)$ such that in c the imperative provides an answer to it, g is any prioritizing ordering source, and speaker and addressee consider g the relevant criteria for resolving $\Delta(c)$;

or else, (ii) in c there is no salient decision problem $\Delta(c)$ such that the imperative provides an answer to it in c , and g is speaker bouletic.

In order to account for advice uses, Kaufmann has to hard-code the fact that the imperative provides a solution to a decision problem into the conventional felicity conditions of imperatives. Because of this, she then is forced to provide the second disjunct as an escape-hatch, in order to be able to account for other uses of imperatives that are not connected to the resolution of decision problems at all (such as well-wishes and absent wishes, for example).

The analysis adopted here leaves reasoning about decision problems entirely in the pragmatics. As a result, the conventional meaning and dynamic effect of imperatives can be greatly simplified.

6.2 Performative uses of desideratives

In Condoravdi and Lauer (2009), we drew attention to the fact that assertions about speaker desires can be used to perform a range of speech acts that is strikingly similar to those that can be performed by means of imperatives:

- (6.40) a. [Mother to child]
I want you to clean your room before playing! (Command)
- b. [Mother to child]
You do NOT want to touch that cookie! (Prohibition/Warning)
- c. [Doctor to patient]
I want you to take these pills for a week. (Advice)

- d. [Recipe]
You want to stir the mixture well. (Advice)
- e. [Affirming an offer]
No, really, I want you to take the last cigarette. (Invitation)
- f. [Among collaborators]
I want you to write this up before our next meeting. (Request)
- g. If it is that important to you, I want you to go. (Concession)

We spelled out an analysis of these uses in much the same way as I did above for imperatives, but without appeal to the crucial notion of commitment: We specified contextual conditions under which the use in question would arise. As an example, a crucial ingredient for deriving **order** uses was the condition in (6.41):

- (6.41) **Speaker Authority over the hearer:** It is commonly assumed that if the hearer believes that the speaker prefers unsettled, hearer-controllable ϕ over not- ϕ , the hearer is socially or institutionally obligated to execute an action that ensures that ϕ .

Besides omitting reference to the notion of *commitment* (or rather, approximating this notion by reference to what is ‘commonly assumed’, i.e., public knowledge), this formulation does not employ the distinction between effective and other preferences—the obligation arising from an **order**-use of a desiderative assertion is tied to the addressee knowing about any desire of the speaker.

As Sam Cummings (p.c.) correctly observed at the time, this predicts that if the speaker is taken to have authority, then any assertion of a speaker desire should count as a command. But this is not correct, the same mother that gives a command with (6.40a) could employ (6.42) only to inform the child that she desires him to clean his room, but leave up to the child if it wants to clean it (picture a mother who firmly believes in anti-authoritarian parenting).

- (6.42) I want you to clean up your room, honey. Nothing would make me happier.

In Condoravdi and Lauer (2009), we took **want** to simply entail the existence of an unspecified desire, which initially seems plausible. But in our work on anankastic conditionals (Condoravdi and Lauer to appear), Cleo Condoravdi and I have since argued that there is reason to think that the meaning of **want** is underspecified, much in the way in which the meaning of modal auxiliaries is usually taken to be underspecified. A similar intuition was expressed by Hare (1968) when considering the contrast between (6.43) and (6.44).⁷

(6.43) If you want to have sugar in your soup, you should ask the waiter.

(6.44) If you want to have sugar in your soup, you should get tested for diabetes.

“Let us consider the meaning of ‘If you want’ in the two cases. In the ‘diabetes’ case, a first approximation would be to say that it means the same as ‘If you, as a matter of psychological fact, have a desire’. I am very much inclined to deny that it means anything like this in the ‘waiter’ case.”

(Hare 1968)

Given the way I have modeled preferences and desires so far, we can assume that **want** depends on a contextual parameter, viz., a preference structure, which may be the agent’s effective preference structure, or it may be one of the underlying preference structures which model psychological desires, inclinations, appetites, etc. We could then say that the ‘performative’ readings arise only when **want** targets the agent’s effective preferences.

To make this work, at least without modifying the set-up of the system so far, however, we need another ‘introspection’ principle for commitment to beliefs and preferences, namely the one in (6.45).

(6.45) If $ep_{a,t}(p) \in PB_a(t, w)$, then $p \in PEP_a(t, w)$.

Compare (6.45) with the principle we already have introduced in Section 5.3.2:

⁷See Levinson (2003), and references therein, for similar observations.

Doing away with PEP?

Given the success with explaining why desiderative assertions (if **want** targets effective preferences) have performative uses like imperatives, it may seem tempting to turn around and simplify the analysis of imperatives, as well, and indeed of the whole system, in the following way.

We get rid of the independent specification of PEP in the models, and instead define the pep-operator as an abbreviation:

$$(6.49) \quad \text{pep}_a(\phi) := \text{pb}_a(\text{ep}_a(\phi))$$

The desired introspection properties follow immediately, both the pb(pep) and the pb(ep) version. And we can simply represent pb as an accessibility relation, without worrying about how to keep the PEPs in sync. Imperatives simply become assertions about effective preference. And our models are simpler, because they contain one fewer operator. What is not to love about this solution?

This solution would, indeed be pleasing in its representational parsimony. But it does lack one essential feature of the analysis that independently represents PEPs. Commitments to effective preferences, like all commitments, are *stable*: They persist until and unless they are rescinded. The same is not true for plain effective preferences (or, realistically speaking, beliefs). Effective preferences change over time, and speakers are aware of this. They might form new underlying preferences, or their evaluation of their relative importance may change. And even if this does not happen, agents will frequently have to revise their effective preferences to maintain consistency (e.g., when they learn that two previously unranked preferences are in fact incompatible). But if effective preferences are not temporally stable, then commitment to preferences, as defined in (6.49), will not be temporally stable, either. This becomes apparent if we put in the omitted temporal parameters:

$$(6.50) \quad \text{pep}_{a,t}(\phi) := \text{pb}_{a,t}(\text{ep}_{a,t}(\phi))$$

Now consider what happens at a later time t' . Given that doxastic commitment is stable, we still have:

$$(6.51) \quad \text{pb}_{a,t'}(\text{ep}_{a,t}(\phi))$$

But that is not the same, nor does it entail (given that eps can change) that

$$(6.52) \quad \text{pep}_{a,t'}(\phi) := \text{pb}_{a,t'}(\text{ep}_{a,t'}(\phi))$$

And we certainly do not want to revise ep to be timeless—speakers can and do speak about the fact that their preferences have changed in the past or might change in the future, and there is no reason to believe that such talk is always about non-effective preferences.

So it seems we cannot do without an independent representation of the notion of ‘commitment to preferences’, next to ‘commitment to belief’, because without one, only the latter notion would be stable—and we need commitment to preferences to be stable if we want to use the notion to explain the workings of predicates like **promise**: A speaker who promises to get Sally from the airport does not just commit to act as though he wants to get her unless he changes his mind. If he fails to show up at the airport, he cannot defend himself by arguing ‘No, no, I did not violate my commitment. I acted according to the preference for being there for a while, but two hours before I had to leave, I changed my mind, and so I did not go.’ Such an argument would be ridiculous—and on the PEP-analysis, we can say why: Because commitment to preferences, like all commitment, is stable. On the $\text{pb}_{a,t}(\text{ep}_{a,t})$ -analysis of preferential commitment, this argument would be perfectly plausible. The agent had the requisite commitment at the time, and he has not since acted in a way that is incompatible with him believing that he did—until he changed his mind. But that is not how commitment works.

Unfortunately, these considerations also cast some doubt on the plausibility of the $\text{pb}(\text{ep})$ -introspection property proposed above. This property does not cause instability of peps , of course, but on reflection, it is not clear how plausible it is. Why should doxastically committing oneself to having a preference at time t create an enduring commitment to a preference for p ? Wouldn’t it be just as natural, if not more natural, to say that such a doxastic commitment is just that: What endures is your commitment to believe that you had the preference in question at t , but

that does not affect your persisting preferences to commitment. Committing to a preference (indefinitely), and committing to believe to have a preference at t are just two different kinds of things—why should the latter give rise to the former?

6.3 Interrogatives

In the present section, I want to sketch, in a very brief fashion, how interrogative sentences can be integrated into the theory of the form–force mapping proposed here.⁹ Of course, interrogatives are a complex topic, so this should be seen as a pointer to future work rather than a fully spelled out proposal.

For concreteness, I assume that the interrogatives of *Sen* (which, due to the propositional nature of the language, are all polar interrogatives), denote a partition of the set of possible worlds, as in Groenendijk and Stokhof (1984):

$$(6.53) \quad \llbracket ?\varphi \rrbracket^{Sen} = \left\{ \llbracket \varphi \rrbracket^{Prop}, \llbracket \neg\varphi \rrbracket^{Prop} \right\}$$

Interrogatives, just like declaratives and imperatives, are associated with a convention of use, the INTERROGATIVE CONVENTION:

$$(6.54) \quad \text{INTERROGATIVE CONVENTION (NL statement)}$$

When a speaker utters an interrogative with content $?\varphi$, he thereby incurs the following commitment:

$$\text{pep}_{Sp}(\exists p \in ?\varphi : \text{pb}_{Addr}(p))$$

That is, with an interrogative $?\varphi$, a speaker commits himself to a preference for the addressee to be committed to one of the possible answers to ϕ . More briefly, we can paraphrase this as: The speaker requests that the addressee assert one of the possible answers to his question. Viewed this way, the proposal bears some similarity to the ‘imperative-assertoric’ approach pioneered by Lewis and Lewis

⁹This section summarizes the preliminary proposal for an analysis of interrogatives in Lauer and Condoravdi (2012).

(1975), which is arguably the analysis with the broadest coverage of uses among the ‘imperative paraphrase’ accounts which were *en vogue* in philosophical logic for in the 1960s and 1970s.¹⁰ According to Lewis and Lewis’ proposal (elaborated somewhat in Åqvist (1983)), an interrogative $?φ$ is equivalent to the imperative ‘Tell me truly whether $[[?φ]]$ ’—and is thus a request for the true answer. The account in the present framework does not involve the notion of truth, and it does not specify that the speaker requests a *speech act*. Instead, he only is committed to a preference for a doxastic commitment. The latter fact makes the present account more promising in view of *rhetorical questions*, such as (6.55).¹¹

- (6.55) [B does not stop complaining about how bad the movie was]
A: Well, who insisted that we see it?

On the present account, we can view rhetorical questions in the way Rohde (2006) analyzed them, as ‘redundant interrogatives’—they are redundant because the commitment that the speaker commits to preferring already exists.¹²

How does this proposal fare in accounting for the other uses of interrogatives? A final answer to this question of course requires careful spelling out of the various conditions that give rise to the various uses, but, at first glance, it does pretty well. Consider the sample of uses in (6.56).

- (6.56) a. Is it raining? (requesting information)
b. What is the formula for sulphuric acid? (testing knowledge)
c. [A is desperately looking for her keys]

¹⁰The most notable exponent of this kind of theory was Åqvist (1965, et seq.).

¹¹It has sometimes been claimed (most recently and most explicitly by Han (2002)) that rhetorical questions always are associated with (or even assert) their negative answers—i.e., **no** for a polar question, **no-one** for a **who**-question, **nothing** for a **what**-question. Examples like (6.55) show that this is not quite right. Rohde (2006) provides more data challenging this generalization.

¹²Rohde also cashes out ‘redundancy’ in terms of prior commitment. She says that there is a ‘felicity condition’ on the use of rhetorical questions ensuring prior commitment. This could mean that she takes rhetorical questions to be associated with this condition by convention—thus assuming that rhetorical questions constitute their own sentence type. In the present proposal, we could simply *define* a rhetorical question as an interrogative that is uttered in the presence of (obvious) pre-existing addressee-commitment (which might also be what Rohde had in mind).

- B: Could they be in the car? (pointing out a possibility)
- d. Senator, should taxes be raised to balance the budget? (combative question)
- e. And doesn't this line bisect each of these spaces? (Socratic question)

Combative questions like (6.56d) come out as somewhat of a base case—they can be taken to be motivated solely by a desire to get the addressee to commit. Information questions like (6.56a) arise straightforwardly in situations in which the speaker lacks information, assumes that the addressee has the information, and further assumes that the addressee would only commit himself in something he believes to be true. Exam questions like (6.56b) similarly do not pose a problem. In this case, the speaker does not ask to acquire information, but rather simply to get the addressee to commit to an answer, so that he can evaluate his knowledge. Socratic questions like (6.56f) can be viewed as cases where the speaker wants the addressee to consider the correct answer to the question, and forces him to do so by requesting a commitment. Similar things can be said about cases like (6.56d).

Again, this is meant only as a short illustration of how an account of interrogatives would fit into the current, commitment-based view of clause-typing. Many of the details need to be spelled out. The suggested analysis holds promise, however, for a successful treatment of interrogatives that covers their varied uses.

6.4 Denotation types and the form–force mapping

In the statement of the conventions of use, and the associated discussion, I have assumed that declaratives and imperatives both denote propositions, while interrogatives denote sets of those. In terms of the semantics of the language *Sen*:

- (6.57) a. $\llbracket \vdash \varphi \rrbracket^{Sen} = \llbracket \varphi \rrbracket^{Prop}$
 b. $\llbracket !\varphi \rrbracket^{Sen} = \llbracket \varphi \rrbracket^{Prop}$
 c. $\llbracket ?\varphi \rrbracket^{Sen} = \left\{ \llbracket \varphi \rrbracket^{Prop}, \llbracket \neg\varphi \rrbracket^{Prop} \right\}$

In the present section, I want to show that these assumptions are non-essential: While the conception of the form–force mapping proposed in this dissertation (and Condoravdi and Lauer 2011 and Condoravdi and Lauer 2012) puts some constraints on denotational types, it by no means requires the types in (6.57). In this section, I will illustrate how the theory is compatible with a wide variety of assumptions about denotational types. In each case, I will not argue in favor of any particular proposal (such arguments would have to rely on considerations external to the theory proposed here). The intention is simply to demonstrate how various proposals for denotation types could be adopted in the current theory without changing its essential nature.

The present section concerns possibilities that maintain the assumption that sentences have the same kind of denotation in their embedded and matrix uses. Section 6.5 will discuss the possibility of lifting this assumption in order to represent force directly in the semantic denotations of sentences, as proposed by Krifka (to appear).

6.4.1 Denotations for imperatives

We start with imperatives. There is great variation in the semantic literature with respect to the question of what imperatives denote, and the choice made in the preceding discussion—a proposition that characterizes the fulfillment conditions of the imperative—is not a popular one.¹³ What I want to show in the sequel is that, in the present set-up, we could adopt any sensible proposal for what imperatives denote. All that adopting a non-propositional denotation requires is a slight modification of the IMPERATIVE CONVENTION, repeated here for the propositional case:¹⁴

¹³The reason for this is not entirely clear to me—one consideration seems to be the pre-theoretical intuition that imperatives, unlike declaratives, do not relate to *truth*, and cannot be used to make claims. But that does not mean their compositionally-determined denotations cannot be propositions. As pointed out in Section 3.2, a proposition (in the linguist’s or logician’s sense—i.e., a set of possible worlds), by itself, is no more closely related to truth and claiming as it is to fulfillment or desiring.

¹⁴Throughout this section, I use a natural language formulation of the conventions, as those are much easier to read than the corresponding constraints on pragmatic models. I spell out the

(6.58) IMPERATIVE CONVENTION (propositional denotation)

A speaker Sp who utters an imperative with content p thereby incurs the following commitment:

$$\text{pep}_{Sp}(p)$$

Properties Hausser (1978) proposes that imperatives denote properties of individuals: “roughly that property which the speaker wants the hearer to acquire” (p. 84). We could easily adopt this analysis, replacing (6.58) with (6.59).

(6.59) IMPERATIVE CONVENTION (property denotation)

A speaker Sp who utters an imperative towards addressee Ad with content P thereby incurs the following commitment:

$$\text{pep}_{Sp}(P(Ad))$$

Predicates restricted to addressees Portner (2005) adopts a variant of Hausser (1978)’s analysis. Motivated by the fact that imperatives can have overt subjects (such as **Everyone go home!**), Portner proposes that imperatives always have subjects, assuming a null subject referring to the addressee(s) when no overt subject is present.¹⁵ He then takes imperatives to denote predicates that are presuppositionally restricted to only apply to the subject. i.e., **Leave!** has the denotation in (6.60) (where c is the context of use):

(6.60) $\lambda x : x = \text{addressee}(c) . x \text{ leaves}$

normative consequences in terms of the operators pb and pep (omitting all temporal parameters), though.

Another possible motivation to give imperatives non-propositional denotations may be the fact imperatives rarely embed, and generally cannot occur as complements to predicates that take declarative complements. If imperatives have a non-propositional denotation, this can be seen as due to a type clash.

¹⁵This assumption is commonly made, see Kaufmann (2012, Chapter3, 105–122) for extensive discussion.

Again, we could adopt this analysis easily. All we need to do is to define a function *def* that takes predicates into the individual(s) for which they are defined.¹⁶ Then our convention would go as follows:

(6.61) IMPERATIVE CONVENTION (restricted predicate denotation)

A speaker *Sp* who utters an imperative towards addressee *Ad* with content *P* thereby incurs the following commitment:

$$\text{pep}_{Sp}(P(\text{def}(P)))$$

Event descriptions A final example I want to mention is the idea that imperatives denote something like *actions* or predicates of actions, or something of the kind (Mastop (2005) is a recent exponent of this idea). This idea is quite intuitive as long as one focuses on the stereotypical uses of imperatives, i.e., directives. In some sense, these are always about actions (cf. Section 6.1.2), even when they involve otherwise non-agentive predicates. It may hence seem plausible to say that such non-agentive predicates somehow get ‘coerced into’ or ‘conceptualized as’ actions in such uses. Functional heterogeneity, however, casts doubt on this idea. The well-wish **Get well soon!** arguably does not involve any kind of action at all. A more defensible variant of this view is to construe the denotation of imperatives as related to *events* (or more generally *eventualities* in view of stative imperatives such as **Please, be at home!**). For concreteness, let us assume that imperatives denote functions of type $\langle ev, t \rangle$, where *ev* is the type of events.

Again, it would be easy to adjust the IMPERATIVE CONVENTION accordingly. Let *Happen* be a predicate of events true if the event is instantiated. Then:

(6.62) IMPERATIVE CONVENTION (properties of events)

A speaker *Sp* who utters an imperative towards addressee *Ad* with content

¹⁶This could be a single (possibly plural) individual, in case we allow for plural individuals, or a set of individuals. The denotation below is formulated in terms of the single-individual case, but it should be obvious how it could be translated to the set-of-individuals case.

P thereby incurs the following commitment:

$$\text{pep}_{sp}(\exists e : \text{Happen}(e) \wedge P(e))$$

Upshot The discussion of these examples illustrates the following fact: All that the theory proposed here requires is that the denotation of imperatives—whatever it is—can be used to *determine* a proposition that characterizes the fulfillment conditions of the imperative. As long as this is possible, the IMPERATIVE CONVENTION needs to be adjusted only very slightly to be compatible with any kind of denotation.

In Section 6.4.4 below, I will argue that there is at least one further option for the denotation of imperatives that leaves the predictions of the current theory unaffected, according to which the denotation of imperatives does not even directly determine fulfillment conditions. Before turning to that question, I briefly want to show that the considerations of the present section apply, *mutatis mutandis*, to interrogative sentences, as well.

6.4.2 Denotations for interrogatives

Just as the hypothesized IMPERATIVE CONVENTION does not, in any essential way, depend on a particular hypothesis concerning the denotational type of imperatives, neither does the INTERROGATIVE CONVENTION proposed in Section 6.3. All that the convention requires is that there is a way to derive, from the denotation, a set of possible answers to the interrogative. The generalized version of the INTERROGATIVE CONVENTION is stated in (6.63):

- (6.63) a. Let Ans be a function that takes the denotation of an interrogative to the corresponding set of propositions characterizing possible answers.¹⁷
- b. INTERROGATIVE CONVENTION (generalized version)
When a speaker utters an interrogative with content $?\varphi$, he thereby

incurs the following commitment:

$$\text{pep}_{Sp}(\exists p \in \text{Ans}(\varphi) : \text{pb}_{Addr}(p))$$

Of course, the influential theory of interrogatives of Groenendijk and Stokhof (1984) (building on Hamblin (1958)) assumes that the denotation of interrogatives just *is* the set of possible answers. In that case *Ans* can just be the identity function, and (6.63b) amounts to the statement in Section 6.3.

But (6.63) is compatible with a large range of alternative hypotheses about interrogative meaning. As an example, take the STRUCTURED MEANING approach to question meaning, which has a long tradition (Hiz (1978) traces it back to Ajdukiewicz (1938)). Krifka (2001a) summarizes the basic idea as follows:

(6.64) Question meanings are functions that, when applied to the meaning of the answer, yield a proposition.

On this approach, answers are not necessarily taken to be propositions, but we still can take *Ans* to return a set of propositions. To take Krifka's example: If **Who read Die Kinder der Finsternis?** denotes the function $\lambda x. \text{READ}(KF)(x)$, we can simply apply this function to all the individuals in the domain to derive the set of possible answers referred to in (6.63).

As in the preceding section, the lesson is that considerations about possible uses of matrix sentences do not, generally, determine a denotation type, they only *constrain* the set of possible types. If we can derive additional constraints on possible types from investigating the semantics of embedded occurrences (as has been done in the case of interrogatives in the large body of literature subsequent to Karttunen (1977)), then these constraints, together with the constraints derived from pragmatic considerations may—but do not need to—determine a unique appropriate denotation type.

¹⁷I leave it as an open question for now how 'possible answer' should be spelled out, so as to allow for both mention-all and mention-some readings of questions.

6.4.3 The (possible) indeterminacy of denotation types

The present theory is compatible with a large range of hypotheses concerning the denotational type of imperatives. At the same time, imperatives can, at least in English, be embedded only to a very limited degree. In particular, there do not seem to be *any* verbal or adjectival predicates that embed imperatives.

One may wonder, then: How can we figure out what the denotational type of imperatives *is*? The short answer is that we cannot. If a sentence type truly never occurs embedded in other sentences, then we only have considerations about possible uses to guide our hypotheses about the denotational type. And such considerations will never single out one single type. But this is no reason to despair. It simply means that the question ‘What do imperatives really, truly, denote?’ is an *uninteresting* question. To the extent that imperatives do not embed, the interesting question to ask is how they are used, and what kind of content must their denotation determine? If we adopt the generalized version of the IMPERATIVE CONVENTION from Section 6.4.1, the answer is: The denotation must determine the fulfillment conditions of the imperatives. There simply is nothing more to be said about the denotation of imperatives.

It is even possible, if imperatives do not embed, that there is no fact of the matter about what the denotation of imperatives is. What this means depends, of course, on what we take to be the determinants of semantic facts, a difficult philosophical question I wish not to go into here. But let us assume, for the sake of the argument, that what determines semantic facts is the linguistic knowledge of competent speakers of the language. Now, consider the following thought experiment.¹⁸ A speech community *C* speaks a language *L*, which has an imperative sentence type such that imperative sentences never occur as constituents in other sentences. All speakers in *C* are perfectly identical in their linguistic knowledge about *L*,¹⁹ with one exception: There are two subgroups of speakers in *C*. On the

¹⁸On some level, this thought experiment is just a Quinean indeterminacy argument.

¹⁹This assumption is of course one of the crucial problems for the thesis that semantic facts are determined by the linguistic knowledge of the speakers of the language, but grant me the assumption for the sake of the thought experiment.

one hand there is the group C_p , which takes imperatives to denote propositions,²⁰ and believes them to be associated with the appropriate convention of use (i.e., the propositional version we have been using in previous sections). On the other hand, there is the group C_e which takes imperatives to denote properties of eventualities, and also believes them to be associated with the corresponding convention of use. The speakers in C_p and C_e could interact with each other smoothly, using imperatives to communicate successfully all day long, without ever realizing that they have incompatible beliefs about the denotational type of imperatives. Nor could an external observer (a linguist studying L , say) ever determine which of the individuals belongs to C_p and which belongs to C_e . In such a community, there simply would be no fact of the matter as to what the denotation type of imperatives is. There would only be a fact of the matter about how imperatives are used, and a fact of the matter about what imperative denotations need to be able to do (i.e., determine fulfillment conditions). To the extent that imperatives in English truly do not embed, I see no reason not to believe that the same is true for speakers of English.

It is not strictly speaking true, or at least not obvious, however, that imperatives never occur embedded in other sentences: English allows for conditional imperatives, as in (6.65a), which at least *could* be construed as conditional declaratives with an imperative consequent. Similarly, Imperatives can occur as the first element in conjunctive sentences (6.65b) and in disjunctive sentences (6.65c).²¹

- (6.65) a. If you get lost, call me!
 b. Mow the lawn and I will give you \$5.
 c. Freeze or I'll shoot.

It is conceivable that investigation of constructions such as (6.65) will yield further

²⁰When I say that speakers in C_p 'take imperative to denote propositions', I refer, of course, to their implicit grammatical knowledge, not to a semantic theory of imperatives they believe in.

²¹All these three constructions have seen considerable attention in the semantics literature, but a consensus on how they should be analyzed has not yet emerged. On the topic of conditional imperatives, see Kaufmann and Schwager (2009), and references therein. On conjunctions and disjunctions of imperatives and declaratives (so-called 'pseudo-imperatives'), see Kaufmann (2012, Chapter 6, p. 221–254), and references therein.

constraints on what the denotational type of imperatives should be.

6.4.4 One convention for declaratives and imperatives?

As indicated at the end of Section 6.4.1, there is another possibility for the denotation of imperatives that leaves the predictions of the proposal of Condoravdi and Lauer (2012) unaffected. This version of the proposal is analogous to the proposal of Kaufmann (2012). Kaufmann takes imperatives to denote propositions, but not propositions that characterize the fulfillment conditions of the imperative. Instead, on her proposal, imperatives are headed by a modal operator *IMP* much like English **must** or **should**, which is analyzed as a graded modal in Kratzer (1981)'s theory of modality.²² Schematically, we can characterize her proposal as in (6.66).

$$(6.66) \quad \llbracket \text{Leave!} \rrbracket = \text{IMP}(\text{you leave})$$

Following Kaufmann's lead, we could assume that imperatives contain an abstract imperative operator, whose truth conditions are given in terms of *pep*:

$$(6.67) \quad \llbracket \text{Leave!} \rrbracket = \text{pep}_{Sp}(\text{you leave})$$

And assume that *IMPERATIVE CONVENTION* is (6.68).

$$(6.68) \quad \text{IMPERATIVE CONVENTION (Kaufmannian version)}$$

When a speaker *Sp* utters an imperative *S* with content ϕ , he thereby incurs the following commitment:

$$\text{pb}_{Sp}(\phi)$$

Given the semantics in (6.67), that would mean that a speaker who utters an imperative incurs the commitment $\text{pb}_{Sp}(\text{pep}_{Sp}(p))$. Given the introspection properties

²²This is motivated, in part, by the fact that modals can be 'used performatively' to create obligations, etc., instead of reporting them (cf. Kamp (1978)). Kaufmann imposes a number of conventional felicity conditions on imperatives that are intended to ensure that imperatives can only be 'used performatively'.

in Section 5.3.2, this entails $\text{pep}_{Sp}(p)$ —that is, (6.68) comes out as equivalent to the original version of the IMPERATIVE CONVENTION.²³

Of course, (6.68) is identical to our original DECLARATIVE CONVENTION, so if we adopt (6.67), we can replace the two conventions with a single one. Given that, under present assumptions, imperatives and declaratives are the only sentence types that have a propositional denotation, we can do so expediently by having the denotation reference the denotational type:

- (6.69) CONVENTION ABOUT SENTENCES WITH PROPOSITIONAL DENOTATIONS
When a speaker Sp utters an sentence S with content ϕ of type $\langle s, t \rangle$, he thereby incurs the following commitment:

$$\text{pb}_{Sp}(\phi)$$

Deciding between this version of the theory and one on which imperatives semantically characterize fulfillment conditions involves a trade-off in complexity. The version here is simpler in that it requires only a single convention where the previous ones needed two, but this simplicity is traded against increased complexity of the denotations of imperatives, as we need to assume an abstract imperative operator.

6.4.5 The form of the clause-typing conventions

In the statement of the various conventions, I so far have glossed over an issue that constitutes another way in which the present theory is very flexible: The issue of clause type identification. The natural language descriptions of the conventions had the form ‘When a speaker utters a declarative/imperatives/interrogative ...’, but I have not said what I take this to mean.

²³In particular, the analysis would still predict that imperative utterances are ‘automatically sincere’: Even though imperatives, on this version of the analysis, create doxastic commitments, the proposition the speaker gets committed to is true as soon the sentence is uttered (cf. the discussion of explicit performatives in Chapter 7). The same is not true for Kaufmann’s analysis, as we point out in Condoravdi and Lauer (2012).

One way to construe this is as an abbreviation of a morpho-syntactic description. This corresponds most directly to how sentence types are treated in *Sen*—each sentence contains an operator that marks it as being of a given syntactic type, which (at least in the case of imperatives and declaratives) is uninterpreted. We can make this construal more explicit:

(6.70) a. DECLARATIVE CONVENTION (syntactic version)

When a speaker Sp utters a sentence that is morpho-syntactically declarative and has the content ϕ , he thereby incurs the following commitment:

$$pb_{Sp}(\phi)$$

b. IMPERATIVE CONVENTION (syntactic version)

When a speaker Sp utters a sentence that is morpho-syntactically imperative and has the content ϕ , he thereby incurs the following commitment:

$$pep_{Sp}(\phi)$$

c. INTERROGATIVE CONVENTION (syntactic version)

When a speaker Sp utters a sentence that is morpho-syntactically interrogative and has the content ϕ , he thereby incurs the following commitment:

$$pep_{Sp}(\exists p \in \phi : pb_{Ad}(p))$$

But this is not the only option. If we assign a unique denotation type to each sentence type, the conventions can just as well be formulated in terms of these types. Suppose, for concreteness that declaratives are of type $\langle s, t \rangle$, imperatives are of type $\langle ev, t \rangle$ and interrogatives are of type $\langle \langle s, t \rangle, t \rangle$. Then the conventions can be specified as follows:

(6.71) a. DECLARATIVE CONVENTION (denotational type version)

When a speaker Sp utters a sentence that has a content ϕ of type $\langle s, t \rangle$,

he thereby incurs the following commitment:

$$\text{pb}_{Sp}(\phi)$$

b. IMPERATIVE CONVENTION (denotational type version)

When a speaker Sp utters a sentence that has a content ϕ of type $\langle ev, t \rangle$, he thereby incurs the following commitment:

$$\text{pep}_{Sp}(\phi)$$

c. INTERROGATIVE CONVENTION (denotational type version)

When a speaker Sp utters a sentence that has a content ϕ of type $\langle \langle s, t \rangle, t \rangle$, he thereby incurs the following commitment:

$$\text{pep}_{Sp}(\exists p \in \phi : \text{pb}_{Ad}(p))$$

There are other options, but the contrast between these two versions may suffice to illustrate the point. Which of these is more attractive will depend largely on two factors: On the one hand, using the denotational type versions requires that each type of sentence that has its own use also has a unique denotational type. As such, it is unsuitable if we have independent reason (e.g., from the investigation of embedded uses) to assume that two sentence types with distinct uses have the same denotational type. On the other hand, the syntactic version is less attractive if we have a class of sentences that is morpho-syntactically heterogeneous, but whose members have a uniform use. In that case, we would be forced to state multiple conventions for this single class (or to use a cumbersome disjunctive morpho-syntactic description in the statement of the convention).

Finally, if we have a class of sentences that is morpho-syntactically heterogeneous, but that shares its denotation type with other sentences that do not have the same use, neither the syntactic, nor the denotation-type versions are attractive. In such a situation, an attractive solution is to opt for semantic representationalism

of the kind advocated in Potts (2005). In Potts' system, natural language sentences are translated into a typed λ -calculus in an entirely compositional fashion.²⁴ The semantic representation language then allows for expressions with the same model-theoretic interpretation to have distinct semantic types (in Potts' case, types for 'at-issue' meanings and 'conventional implicature' meanings).

Which version of the theory of clause-typing presented here is most desirable thus depends on general considerations of parsimony and theoretical elegance, as many theoretical choices do. But it also depends on what independent constraints we have on denotational types, and on the morpho-syntactic analysis of sentences that share a single use. It hence is not a question that can be resolved just in terms of considerations about use alone.

6.4.6 Conclusion

The purpose of this section was to illustrate that the theory of clause typing proposed here and in Condoravdi and Lauer 2011 and Condoravdi and Lauer 2012 is compatible with a wide range of assumptions about the denotational semantics of sentences of various types. Even though the rest of this dissertation makes particular assumptions about denotational types (declaratives and imperatives denote sets of possible worlds, interrogatives denote sets of those), the overall theory does not rely on these assumptions in any essential way.

I consider this a significant virtue of the theory. Considerations about possible uses do not determine a unique denotatum for sentences, they only contribute some general constraints on what a sentence can denote. Accordingly, the architecture of a general theory of use, including a theory of conventional constraints on use, should not require any particular kind of semantic denotatum, but should only determine some constraints on them. The theory proposed here does just that.

²⁴That is, unlike in strongly representational theories such as DRT, the semantic representation maintains the structure of the interpreted phrase structure.

6.5 Representing force in the compositional system

In this section, I want to take up the question whether sentential force can and should be represented within the system of semantic composition. Up to now we have operated under the assumption that it is not. Matrix sentences have the same denotation as embedded sentences have, and these denotations capture only the *content* of the sentence, not its force.

This is not the only possibility, however. As mentioned in Section 3.3, an alternative is to assume that matrix sentences have a denotation which encodes their sentential force. The most straightforward way to implement this is to assume that matrix sentences contain an unpronounced ‘force operator’ whose semantics specifies the force of the sentence. On the conception of sentential force proposed in the previous chapters, this force consists in *essential normative effect* of the sentence.

The most obvious way to implement this is to assume force operators denote *update functions* that take the current context of utterance into a new context of utterance. In order to implement this smoothly, it would be expedient to slightly adjust our models in the following way: Instead of associating public beliefs and preferences directly with world/time pairs, we would associate each world/time pair with suitable representation c of a *context*, which in turn determines the sets $c(PB_i)$ and $c(PEP_i)$ for each discourse participant i , as well as other features of the content, such as $c(Sp)$ for the speaker of the current utterance. Then the semantics of the force operators would be as follows, where $c[x/y]$ is the context that is like c , except that the component x has the value y :

- (6.72) a. Declarative operator

$$\llbracket \text{!} \rrbracket = \lambda\varphi.\lambda c.c[\text{PB}_{c(Sp)}/c(\text{PB}_{c(Sp)}) \cup \{\varphi\}]$$
- b. Imperative operator

$$\llbracket \text{!} \rrbracket = \lambda\varphi.\lambda c.c[\text{PEP}_{c(Sp)}/c(\text{PEP}_{c(Sp)}) \cup \{\varphi\}]$$
- c. Interrogative operator

$$\llbracket \text{?} \rrbracket = \lambda\varphi.\lambda c.c[\text{PEP}_{c(Sp)}/c(\text{PEP}_{c(Sp)}) \cup \{\exists p \in \varphi : \text{pb}_{c(Ad)}(p)\}]$$

This treatment would essentially be that of Krifka (to appear). Krifka's representation is not completely identical of course, but it is very similar in the essential aspects: He also uses a forward-branching model of time, and he also assumes that utterances reduce the set of future possibilities (he calls this an 'option space'), i.e., for him, too, utterances exclude certain future states of affairs. The commitments that he assumes are different from the ones argued for in the preceding chapters,²⁵ but the basic idea is the same.

Even if we assume that sentences have such dynamic denotations, we still need to assume a convention of use—again represented as a constraint on admissible models—that ensures that if a sentence denotes a certain update function, then an utterance of the sentence applies this function to the current context. Let $con(w, t)$ refer to the context at w, t and $*$ a metavariable ranging over $\{\vdash, ?, !\}$.²⁶

(6.73) UTTERANCE EFFECT constraint

A model for \mathcal{P}_{Sen} is admissible only if for all $*\varphi, t, a, b$:

If $Hap_w(t, t + 1) = utter(a, b, \varphi)$ then

$$con(w, t + 1) = \llbracket *\varphi \rrbracket (con(w, t))$$

As discussed in Section 3.3, even though (6.73) is uniform in that it represents a single convention that governs clauses of all types, it is not at all obvious that this setup is more parsimonious than one that employs multiple conventions of use. After all, we now need a convention of use *plus* one operator per sentence type. And we had to give up the attractive, simple picture according to which the denotation of a sentence is the same in its matrix and embedded uses, assuming a number of silent operators that are otherwise unmotivated.

Krifka discusses a number of ways in which we can exploit a compositional representation of force in the treatment of various phenomena. Here I will focus

²⁵In particular, Krifka's imperative operator directly imposes an obligation on the addressee, which is quite problematic, given the functional heterogeneity of imperatives discussed above.

²⁶Krifka defines a second update operation, $+$ that essentially does the job of this convention of use—so despite the fact that he has dynamic denotations, he too needs an additional operation that applies the denotation to the current context when an utterance happens.

on the phenomenon that I think has the potential to make the most compelling argument for representing force in the compositional system: illocutionary modifiers. In Chapter 7, I will discuss another of Krifka's cases, which involves logical operators in explicit performatives.

Illocutionary modifiers are items like **frankly** which intuitively predicate something of the current speech act.

(6.74) Your tie and shirt frankly don't go together.

Krifka proposes that **frankly** combines with the force operator in order to modify it, but **frankly** can be treated similarly on the assumption that force is extracompositional. We can assume that **frankly** indexically refers to the current utterance, and predicates something of it (namely that it is done frankly). All that remains to be explained, then, is why the contribution of **frankly** does not become part of the truth conditional content of the sentence—but that could be accounted for in various ways, e.g., in a multi-dimensional system à la Potts (2005).

The reason this works smoothly in the case of **frankly** is that its contribution leaves the conventional dynamic effect of the sentence unaffected: A declarative utterance with **frankly** commits the speaker to believe in the truth of the content of the sentence, just as an utterance of the same sentence without **frankly** would.

Krifka does not discuss other such modifiers, but it is quite possible that there are ones that work differently in that they *change* the basic dynamic effect of the uttered sentence. If such operators exist, they would constitute a compelling argument for the representation of force in the compositional system.

A potential example are the 'illocutionary evidentials' Faller (2002) describes in Cuzco Quechua. According to Faller, declaratives containing the hearsay evidential **-si** do not create speaker commitments. If this is so, we cannot treat **-si** in the way I just suggested for **frankly**, as **-si** needs to *cancel* the conventional effect of the declarative—putting **-si** into a sentence prevents the usual commitment from coming into effect.²⁷

²⁷Some particles in languages like German, in particular the ones that have been called the

For items like evidentials, which typically form a small closed class of morphemes which are often in complementary distribution, another treatment naturally suggests itself. We could see them as marking distinct sentence types—i.e., we could assume that Cuzco Quechua does not have one declarative, but several, each of which is associated with its own extra-compositional convention of use. On such a construal, there would be no commitment for **-si** to cancel—instead, **-si** would function simply as a marker for the ‘hearsay-declarative’, much like word order and intonation function as markers for sentence types in English.

It is an open question whether there are expressions that (i) do not affect the truth-conditional content of sentences, (ii) affect the commitment undertaken with a sentence and (iii) for which it is implausible to assume that they serve to mark their own clause type. If there are such expressions, their existence constitutes a compelling argument for a compositional implementation of clause-typing à la Krifka. But unless and until we can make such a compelling case, I prefer to keep the specification of normative effects of utterances out of the denotational semantics of natural language sentences.

Abtönungspartikel (‘downtoning particles’) might constitute a similar case—for example, Zimmermann (2004) proposes that **wohl** weakens that degree of commitment that the speaker undertakes with his utterance.

Chapter 7

Explicit performatives

Explicitly performative utterances (explicit performatives, for short) are utterances of sentences like those in (7.1a-c). They are of interest because they are a rare instance where ‘saying makes it so’: If a speaker sincerely *says* that he promises, he has thereby promised.

- (7.1)
- a. I promise to come to the party.
 - b. I order you to sign the contract.
 - c. I claim that the king is illegitimate.

Explicit performatives are particularly interesting from the perspective of a theory of clause-typing, as they apparently are an instance where illocutionary force of an utterance is made explicit. This is how explicit performatives have been typically conceptualized since the seminal work of Austin (1962).

One might hence suspect that the theory of clause-typing presented in the previous chapters faces difficulties in accounting for explicit performatives, given that it intentionally avoids any reference to illocutionary concepts. In fact, the opposite is the case: The analysis of explicit performatives is entirely straightforward.

7.1 Saying makes it so

To appreciate how surprising it is, initially, that speakers can do something (such as promising) by saying that they do, it is useful to consider essentially non-linguistic acts, such as frying an egg.

(7.2) John: I (#hereby) fry an egg.

By uttering (7.2), John does not bring about the frying of an egg. We would not report the utterance in (7.2) by saying **John fried an egg**. Instead, we would report it, for example, by saying **John said he fried an egg**.

Things are different with explicit performatives. If John says **I promise to come to the party**, we can report his utterance by saying **John promised to come to the party** (as well as the more prolix **John said he promises to come to the party**). This is a general property of explicit performatives: We can report utterances of the (a)-sentences in (7.3)–(7.5) by using the corresponding (b)-sentence.

(7.3) a. I promise to come to the party.
b. He promised to come to the party.

(7.4) a. I order you to sign the contract.
b. He ordered me to sign the contract.

(7.5) a. I claim that the king is illegitimate.
b. He claimed that the king is illegitimate.

There is, of course, an obvious difference between **fry an egg** and **promise to *p***: The latter refers to an action that can be performed by a linguistic utterance, whereas the former cannot. But this cannot be the only difference between the two predicates that is responsible for the fact that, with performatives, saying makes it so. There are many predicates that refer to communicative actions which cannot be ‘used performatively’—i.e., which cannot be used to perform the action merely by saying that one does.

- (7.6) a. I (hereby) insult you.
 b. I (hereby) annoy you.
 c. I (hereby) intrigue you.
 d. I (hereby) frighten you.

Insulting, annoying, intriguing and frightening are certainly ‘things that can be done with words’ (to use Austin’s phrase). And yet, these are not things a speaker can do merely by saying that he does.¹ Explicit performatives, then, raise two questions:

1. How can, with certain verbs, saying that one does something constitute doing that thing?
2. What is the difference between verbs that can be used in this way and verbs that cannot?

The traditional way of approaching this question answers the second question first, and then uses this answer to tackle the first: A class of illocutionary *acts* is characterized independently, and distinguished from another class of acts that can be performed by making an utterance but are not illocutionary acts (these are called, following Austin, the *PERLOCUTIONARY* acts). It is then hypothesized that verbs like **promise**, **order** and **claim** that can be used in explicit performatives refer to illocutionary acts, while verbs that cannot be so used, like **insult**, **annoy** and **frighten** instead refer to perlocutionary acts. This hypothesis is then used in trying to explain why illocutionary verbs can be used in explicit performatives.

The strategy here will be to go the other way round: I will specify suitable meanings for some of the ‘illocutionary’ verbs, and show how it is that they can be used in explicit performatives. Then I will step back and ask, given the nature of this explanation, what a verb has to be like in order to fit this use, in effect

¹With the exception, perhaps, of certain, very special contexts—e.g., if it is clear that any further word out of John’s mouth would annoy Mary, then he can indeed annoy her by saying **I annoy you**—but he also could say anything else. In such a special context, there is nothing about the *content* of the uttered sentence that makes it achieve what it does.

reconstructing the distinction between ‘illocutionary verbs’ and ‘perlocutionary verbs’ in terms of the properties necessary for explicit performatives.

7.2 Explicit performatives as ‘self-verifying’

Approaches to explicit performatives come in two broad classes, and while I defer comparison with selected existing approaches to Section 7.6, I want to briefly situate the analysis that flows naturally from the system of clause-typing developed in the preceding chapters with respect to these classes.

The distinction between the two classes is perhaps most easily appreciated by considering the characterization of explicit performatives offered by Heal (1974):

“Roughly speaking, an explicit performative utterance occurs when

- (i) a sentence is uttered and an action is thereby performed, and
- (ii) the grammatical form of the sentence makes it look at first glance as though the speaker states that he performs that action.”

(Heal 1974, p. 106)

The two classes of accounts that I want to distinguish differ on whether they take the impression described in (ii) to be misleading or not.²

Austin (1962) firmly is in the former camp—he takes it to be obvious that, promising by saying **I (hereby) promise to *p*** does not involve stating that one is promising:³

“In these examples it seems clear that to utter the sentence (in, of course, the appropriate circumstances) is not to *describe* my doing of what I should be said in so uttering to be doing or to state that I am doing it:

²Heal herself argues for the position adopted here—that the impression is *not* mistaken. She does not spell out how (i) is to be explained, though.

³Technically, these remarks by Austin pertain to examples like **I name this ship the *Queen Elizabeth***—but Austin held that these are equivalent, in the relevant respects, to acts of promising, etc.

it is to do it. None of the utterances cited is either true or false: I assert this as obvious and do not argue it. It needs argument no more than that ‘damn’ is not true or false[.]”

(Austin 1962, p. 6)

However, in the years since Austin said this, a large number of authors have held the opposite—that, just as it appears, when a speaker utters an explicit performative, he states that he is doing something and *thereby* does it (Lemmon 1962, Hedenius 1963, Heal 1974, Bach and Harnish 1979, Ginet 1979, Bierwisch 1980, Leech 1983, a. o.).

Though these approaches differ in the details, the idea is to analyze explicit performatives on the model of sentences like (7.7).

(7.7) I am speaking now.

On standard assumptions, (7.7) expresses the proposition that the speaker is speaking at the time of utterance. As a consequence, whenever the sentence is uttered, the proposition it expresses is thereby made true. The idea is to assume that explicitly performative utterances are ‘self-verifying’ in just this way.⁴

Such approaches can be called ‘assertoric’, since they construe the utterance of explicit performatives as being an assertion which makes its content true. If such an approach can be made to work, it is intuitively appealing and parsimonious: The special ‘performative force’ of explicit performatives comes about only in virtue of the fact that utterances of the sentence happen to make the sentence true. And given that explicit performatives look and behave syntactically like other declarative sentences in every way, there is no grammatical basis for the

⁴To say that explicitly performative utterances are self-verifying is, of course, to speak somewhat sloppily—or at least involves some unstated assumptions. An utterance can only be self-verifying—make itself true—if utterances can be true or false, which is a non-trivial assumption. We do not need to assume that utterances (rather than uttered sentences) have truth-conditions in order to make sense of such talk: We can also just take talk of the truth or falsity of a (declarative) utterance to metonymically refer to the truth or falsity of the uttered sentence. In any case, when, here and in the following, I speak of explicit performatives being self-verifying, I mean this in the sense of Lemmon (1962)’s title—they involve ‘sentences that are verifiable by their use’.

once popular claim (initiated by Ross (1970)) that in explicit promises, **I promise** spells out an illocutionary operator that is silent in sentences that are not explicit performatives. Such an analysis, in any case, leaves unanswered the question of how force is related to compositional meaning and, consequently, does not explain how the first person and present tense are special, so that first-person present tense forms can spell out performative prefixes, while others cannot. Minimal variations in person or tense remove the ‘performative effect’:

(7.8) I promised you to be there at five. (is not a promise)

(7.9) He promises to be there at five. (is not a promise)

Explicit performatives look like other declarative sentences, so the most straightforward assumption is that they *are* like all other declaratives in that they have denotations that can be true and false; and that they are used in just the way declarative sentences are normally used, i.e., to make statements or claims. On the analysis advocated in this dissertation, this means that explicit performatives, like all other declaratives, commit their speaker to their compositionally determined content.

We can give precise formal content to what we want to explain. The idea is that an explicitly performative utterance is automatically true once it is uttered, that is, we are looking for a semantic analysis of the verbs involved that ensures the following:

(7.10) If a speaker utters **I promise** ϕ in context c and world w , then $w \in \llbracket \text{I promise } \phi \rrbracket^c$.

(7.11) If a speaker utters **I order** ϕ in context c and world w , then $w \in \llbracket \text{I order } \phi \rrbracket^c$.

(7.12) If a speaker utters **I claim** ϕ in context c and world w , then $w \in \llbracket \text{I claim } \phi \rrbracket^c$.

7.3 Truth conditions for performative verbs

If we assume that explicit performatives are just utterances of regular declaratives, and that all their constituents have the same meaning they would have in other sentences, a good strategy is to first investigate *reportative* uses of performative verbs, in order to arrive at a hypothesis for their meaning. On the one hand, we ultimately require an analysis of these verbs that is adequate for both performative and reportative uses, as Szabolcsi (1982) stressed:

“Interestingly enough, researchers who find the felicity conditions of ‘I promise that...’ or ‘I congratulate you’ so complex and necessary to study closer never devote much attention to the “special truth conditions” of ‘Peter is congratulating Mary’ or ‘Yesterday Peter promised that...’, although it is inevitable that in deciding whether Peter is in the set of those who are congratulating Mary etc. one should consider the very same questions.”

(Szabolcsi 1982, p. 530)

On the other hand, while native speakers have intuitions about the truth conditions of reportative uses, it is unlikely that they have reliable intuitions about performative uses. The reason for this is that, by assumption, performative sentences are automatically true once uttered, and so speakers are unlikely to have intuitions about conditions under which they would be false.

Let us consider, then, the question what has to be the case for (7.13) to be true.

(7.13) John promised to be at the airport at noon.

(7.13) is true if there was an utterance⁵ by John, and with that utterance, John incurred a certain commitment. What kind of commitment? To comply with a promise to *p* is to choose one’s actions so as to make *p* true—that is, to choose

⁵Or some other communicative act. For simplicity, I will talk about ‘utterances’.

one's actions as if one effectively prefers that p . So (7.13) says that John committed himself to pep_{John} (John be at the airport at noon).⁶

What about (7.14)?

(7.14) John ordered Mary to be at the airport at noon.

Again, to comply with an order (or a request, or a similar directive) is to act as though one preferred what has been ordered. One need not actually prefer it, but one must choose one's actions as if one did.

So we can understand (7.15) as saying that there was an utterance by John (to Mary), and that utterance publicized John's (effective) preference for Mary to act as though she effectively prefers to be at the airport at noon. In terms of our formalization of these attitudes, the commitment resulting from John's utterance is $\text{pep}_{\text{John}}(\text{ep}_{\text{Mary}}(\text{Mary be at the airport at noon}))$.⁷

Finally, (7.15) is intuitively true if John made an utterance that committed him to a belief that the king is illegitimate:

(7.15) John claimed that the king is illegitimate.

I summarize the hypothesized semantics of our three performative predicates in (7.16).

⁶**Promise** has some additional presuppositions. Firstly, (7.13) presupposes that John took it for granted, at the time of the promising, that his addressee had a stake in whether or not he will be at the airport at noon. When we think of promises, we usually think of cases where the promiser takes the addressee to desire what is promised, but a sentence such as **I promise that p** can also be used to threaten—i.e., to commit to acting in a way that the promisee disprefers. Crucially, such utterances can still be reported using the verb **promise**: **He promised to make my life miserable**. Secondly, **promise** likely presupposes that the promiser believes that he is able to ensure the truth of what is promised.

⁷**Order** again comes with additional presuppositions. One of these is that the orderer presumes to have authority over the addressee with respect to order, in the sense explicated in (6.22) in Section 6.1.2. *Nota bene*: **order** presupposes only that the speaker presumed to have authority, not that he actually did, as attested by the coherence of (i).

(i) John ordered Mary to be at the airport at noon, but he did not have the authority to do so.

- (7.16) a. **John claimed that p** is true iff John made an utterance resulting in $\text{pb}_{\text{John}}(p)$.
- b. **John promised to p** is true iff John made an utterance resulting in $\text{pep}_{\text{John}}(p)$.
- c. **John ordered Mary to p** is true iff John made an utterance resulting in $\text{pep}_{\text{John}}(\text{ep}_{\text{Mary}}(p))$

What all three of these have in common, of course, is that the meaning of the performative predicates all are spelled out in terms of the normative consequences of utterances. This is, of course no accident: As we will see in the following, it is precisely this property that enables **claim**, **promise** and **order** to work in explicit performatives.

7.4 Deriving self-verification

With the hypothesized meanings for performative predicates, we have in place everything that is necessary to derive the ‘performative effect’: First-person, present tense sentences with these predicates will be self-verifying. Informally, the reasoning goes as follows. Suppose John utters (7.17), and call this utterance u^* .

- (7.17) I promise to be at the party.

The denotation of (7.17), in this context, can be paraphrased as in (7.18):

- (7.18) There is an utterance event u by John that commits him to $\text{pep}_{\text{John}}(\text{John be at the party})$.

(7.17) is a declarative, so by the declarative convention, John’s utterance commits him to the belief that the content obtains:

- (7.19) As a result of u^* , John is committed to believe that there is an utterance u that commits him to $\text{pep}_{\text{John}}(\text{John be at the party})$.

But this, in turn, means that (if commitment to belief is closed under logical consequence):

(7.20) As a result of u^* , John is committed to believe that he is committed to pep_{John} (John be at the party).

But being committed to believing one is committed to a preference means that one is also committed to the preference.⁸ And so:

(7.21) As a result of u^* , John is committed to pep_{John} (John be at the party).

But now, reconsider the content of John's utterance.

(7.22) There is an utterance event u by John that commits him to pep_{John} (John be at the party). =(7.18)

u^* , the utterance of **I promise to be at the party** itself, serves as a witness for the existential claim made by the sentence. But that means that the sentence is self-verifying in the appropriate sense.

Given the formal set-up in Chapter 5, we need not leave things on this informal level: If we extend the object language of our dynamic pragmatics with predicates like **promise**, we can *prove* that these are self-verifying.

7.4.1 Introducing illocutionary verbs into *Sen*

We enrich *Sen* with a set of operators that model performative verbs:

(7.23) For a given \mathcal{P}_{Sen} -model, *Perf* is the smallest language that contains *Sen* and all sentences of the form $P(a, \varphi)$, for $I(a) \in \text{Ag}$, $\varphi \in \text{Sen}$ and $P \in \{\text{claim, promise, order}\}$, as well as propositional combinations of these.

⁸With Condoravdi and Lauer (2011), I take this to be an obviously plausible desideratum for any account of commitments to beliefs and preferences. Formally, this is captured in the introspection properties required in Section 5.3.2.

For the sake of simplicity, I introduce *claim*, *promise* and *order* as sentence-forming operators. Ultimately, we would want to treat them as predicates in a first-order language with quantification over individuals (and times, which we ignore here), but for present purposes, treating them as sentence-forming operators is sufficient.

The denotation of *Sen*-sentences is as before. The new formulas of *Perf* are interpreted as follows:⁹

$$(7.24) \quad \begin{aligned} \text{a. } \llbracket \text{claim}(a, \varphi) \rrbracket &= \{w \in W \mid w \vDash \exists u : \text{utter}_u(a, \varphi) \wedge \text{Result}_u(\text{pb}_a(\varphi))\} \\ \text{b. } \llbracket \text{promise}(a, \varphi) \rrbracket &= \{w \in W \mid w \vDash \exists u : \text{utter}_u(a, \varphi) \wedge \text{Result}_u(\text{pep}_a(\varphi))\} \\ \text{c. } \llbracket \text{order}(a, b, \varphi) \rrbracket &= \\ &\{w \in W \mid w \vDash \exists u : \text{utter}_u(a, b, \varphi) \wedge \text{Result}_u(\text{pep}_a(\text{ep}_b(\varphi)))\} \end{aligned}$$

As in the informal paraphrases, the denotation of *claim*, *promise* and *order* predicate the existence of an *event token* (i.e., that an utterance takes place), which has certain normative consequences.

7.4.2 Performatives with promise and order are self-verifying

We can now represent a sentence like (7.25) as in (7.26) (assuming that *a* is the speaker of (7.25)):

(7.25) I promise φ .

(7.26) $\text{promise}(a, \varphi)$

And we can show that (7.26) is indeed self-verifying, i.e., that:¹⁰

Fact 1. *If $w \vDash \text{utter}_t(u, a, \ulcorner \text{promise}(a, \varphi) \urcorner)$ then $w \vDash \text{promise}(a, \varphi)$.*

⁹Note that *Perf* is untensed, and that formulas like $\text{claim}(a, \ulcorner \varphi \urcorner)$ are hence timelessly true or false, just like the propositional atoms of the underlying propositional language. To reduce notational clutter, I omit the temporal parameters of *pb* and *pep* throughout. In definition in (7.24), the appropriate temporal parameter would obviously be the first component of the interpretation of *u*.

¹⁰Here and in the following, I occasionally enclose *Perf*-formulas in $\ulcorner \cdot \urcorner$ when they occur as arguments to predicates, to make the formulas easier to parse.

- (7.27) a. $w \models \text{utter}(u, a, \ulcorner \text{promise}(a, \varphi) \urcorner)$ (assumption)
 b. $w \models \text{Result}_u(\text{pb}_a \ulcorner \text{promise}(a, \varphi) \urcorner)$ (a) + DECL. CONVENTION
 c. $\text{promise}(a, \varphi) \models \exists u' : \text{Result}_{u'}(\text{pep}_a(\varphi))$ (def promise)
 d. $\exists u' : \text{Result}_{u'}(\text{pep}_a(\varphi)) \models \text{pep}_a \varphi$ (Result factive)
 e. $\text{promise}(a, \varphi) \models \text{pep}_a \varphi$ (b) and (c)
 f. $w \models \text{Result}_u(\text{pb}_a(\text{pep}_a(\varphi)))$ (Result closed)
 g. $\text{pb}_a(\text{pep}_a(\varphi)) \models \text{pep}_a(\varphi)$ (pb(pep) introspection)
 h. $w \models \text{Result}_u(\text{pep}_a(\varphi))$ (Result closed)
 i. $w \models \text{utter}(u, a)$ (a)
 j. $w \models \text{utter}(u, a) \wedge \text{Result}_u(\text{pep}_a(\varphi))$ (i) + (h)
 k. $w \models \text{promise}(a, \varphi)$ (j)+ def promise

Step (7.27g) is the crucial step where we exploit the closure property of (5.32c) of Section 5.3.2, corresponding to the principle **Doxastic reduction for preference commitment** in Condoravdi and Lauer (2011, p. 156). We immediately obtain the same result for **order**:

Fact 2. *If $w \models \text{utter}_t(u, a, b, \ulcorner \text{order}(a, b, \varphi') \urcorner)$ then $w \models \text{order}(a, b, \varphi')$.*

The proof proceeds exactly as in (7.27), with $\text{ep}_b(\varphi')$ in place of φ .

7.4.3 Performatives with claim are self-verifying

The proof is analogous for claim, but I spell it out in full, seeing as it relies on a different introspection principle (viz., Condoravdi and Lauer (2011)'s **Positive introspection for doxastic commitment**, captured here in (5.34a) in Section 5.3.2).

We again can represent (7.28) as (7.29) (assuming a is the speaker of (7.28)):

(7.28) I claim that φ .

(7.29) $\text{claim}(a, \varphi)$

And then show:

Fact 3. *If $w \models \text{utter}_u(a, \ulcorner \text{claim}(a, \varphi) \urcorner)$ then $w \models \text{claim}(a, \varphi)$.*

- (7.30)
- a. $w \models \text{utter}(u, a, \ulcorner \text{claim}(a, \varphi) \urcorner)$ (assumption)
 - b. $w \models \text{Result}_u(\text{pb}_a \ulcorner \text{claim}(a, \varphi) \urcorner)$ (a) + DECL. CONVENTION
 - c. $\text{claim}(a, \varphi) \models \exists u' : \text{Result}_{u'}(\text{pb}_a(\varphi))$ (def claim)
 - d. $\exists u' : \text{Result}_{u'}(\text{pb}_a(\varphi)) \models \text{pb}_a \varphi$ (Result factive)
 - e. $\text{claim}(a, \varphi) \models \text{pb}_a \varphi$ (b) and (c)
 - f. $w \models \text{Result}_u(\text{pb}_a(\text{pb}_a(\varphi)))$ (Result closed)
 - g. $\text{pb}_a(\text{pb}_a(\varphi)) \models \text{pb}_a(\varphi)$ (pb introspection)
 - h. $w \models \text{Result}_u(\text{pb}_a(\varphi))$ (Result closed)
 - i. $w \models \text{utter}(u, a)$ (a)
 - j. $w \models \text{utter}(u, a) \wedge \text{Result}_u(\text{pb}_a(\varphi))$ (i) + (h)
 - k. $w \models \text{claim}(a, \varphi)$ (j)+ def claim

7.4.4 Summary

In foregoing sections, I showed that, given an appropriate specification of the truth-conditional content of performative verbs, the self-verification of explicit performatives immediately results from the system set up in Chapter 5. Even though spelling out the proof is somewhat involved, the treatment of performatives is ultimately extremely simple: Performative verbs have regular, truth-conditional meanings, which happen to be of such a kind that when a performative sentence is uttered, this utterance itself ensures the truth of the sentence.

It is important to stress that the steps laid out in (7.27) and (7.30) are *not* intended as a representation of pragmatic reasoning the hearer goes through when observing an utterance. They are simply proofs that show that self-verification is ensured. Of course, this also means that when a hearer observes an utterance of an explicit performative, he thereby also comes to believe that the uttered sentence is true, i.e., that the promise, order, claim, etc. indeed happened.

7.5 Further predictions

In this section, I draw out some further predictions of the account of explicit performatives presented here. In Section 7.5.1, I briefly discuss how self-verification fails when performative verbs are used reportatively. In Section 7.5.2, I spell out some predictions of the account with respect to combination of verbs with logical operators, showing that the observations that Krifka (to appear) explains in terms of illocutionary operators follow straightforwardly without assuming those. Finally, in Section 7.5.3, I return to the question why verbs like **insult** and **annoy** cannot be used in explicit performatives.

7.5.1 Reportative uses of performative predicates

I have shown that, with first person subjects, sentences involving performative verbs like **promise** are self-verifying. Of course, we want to also assure ourselves that with *non*-first person subjects, self-verification does not obtain. That is, a sentence like (7.31) should not be self-verifying:

(7.31) Mary: John promises to be at the party.

This is indeed the case. It is possible that $w \models utter_u(a, \ulcorner promise(b, \varphi) \urcorner)$, but $w \not\models promise(b, \varphi)$, because a 's utterance event u cannot serve as a witness for the truth of the sentence she asserts, for two reasons. Firstly, $promise(b, \varphi)$ asserts that there is an utterance event *by* b , while u is an utterance by a . Secondly, the proof of self-verification fails at step (7.27g):

(7.32) $pb_a(pep_b(\varphi)) \not\models pep_b(\varphi)$

There is no introspection property that ensures that $pb_a(pep_b(\varphi))$ entails $pep_b(\varphi)$, and there should not be one: Just because I am committed to believe you are committed to prefer something does not mean that you are so committed.

An analogous argument can be made regarding tense, once we extend our object language to represent it. Tense would put restrictions on the temporal location of

the utterance event, and non-present tense would again ensure that the utterance cannot serve as a witness for the truth of the asserted sentence, explaining why (7.33) cannot be used performatively.

(7.33) I promised that I will be at the party.

In general, for self-verification to obtain, the utterance of a sentence must be suitable for serving as a witness for the existential claim the sentence makes. If it does not, the sentence can only be used reportatively.

7.5.2 Performatives and logical operators

Krifka (to appear) makes a number of observations concerning the interaction of logical operators with explicit performatives, which he proposes to explain by hypothesizing non-Boolean meanings for these operators which apply to illocutionary operators that are part of the compositionally-determined denotation of sentences (cf. Section 6.5). In this section, I show that the account of explicit performatives proposed above and in Condoravdi and Lauer (2011) directly predicts these observations on the assumption that the logical operators have their traditional Boolean denotations.

Explicit performatives and negation

The first operator Krifka discusses is negation:

(7.34) I do not promise to come.

Krifka characterizes such sentences as follows:

“Denegation can be understood as explicitly refraining from performing a speech act, like an act of promise (cf. Hare 1970). As such, they are not regular speech acts. This is reflected by the fact that we cannot use **hereby**, different from usual explicit performatives.”

(Krifka to appear)

Krifka then proceeds to propose that **not** in (7.34) does not denote regular Boolean negation, but instead is an operator that applies to an illocutionary operator to yield a new one. But is unclear why that would be necessary: It seems we can explain everything that there is to explain about (7.34), maintaining the assumption that **not** is regular, Boolean negation. On the analysis given here, (7.34) commits the speaker to the belief that he does not promise. This seems sufficient to capture the impression that its speaker is ‘explicitly refraining from promising’: Just like saying nothing or making an alternative, unrelated utterance, (7.34) does not commit the speaker to act as though he prefers to come, but it does so in a particular way: By saying what he does, the speaker is addressing the issue of whether he is promising, thus either making this issue salient or responding to the pre-existing salience of this issue. Is there anything about (7.34) that cannot be explained by the fact that its speaker avoids committing himself to come, and makes the fact that he does so salient and explicit?¹¹

Krifka *appears* to think there is something more to be explained. His analysis predicts that an utterance of (7.34) precludes any future promise to come. But this prediction strikes me as far too strong. With (7.34), the speaker does not take on a commitment that he will not make such a promise in the future, he simply explicitly withholds this promise in the present. This is the difference between (7.34) and (7.35).

(7.35) I will not promise to come.

Krifka does not seem to share this intuition of a difference between the two sentences (“We can execute the same change of the option space by expressing a future assertion, **I will not promise to come.**”), but it seems plain to me that with (7.35), the speaker commits himself to not promising in the future while with (7.34), he

¹¹Hare (1970), which Krifka cites, does not provide any arguments for why (7.34) should not be viewed as simply the assertion of a negated proposition, though he likely would endorse such a view. All that Hare (1970) argues is that even if we regard **I promise** as a ‘performative prefix’ (a term that he does not use, of course), we can deal with such negated examples by assuming that there is such a thing as illocutionary negation. But that is not an argument for the existence of such a negation. It appears that illocutionary negation solves a problem that does not exist in the first place unless we assume illocutionary operators in the semantic representation.

does not. I leave this issue for future research.

Explicit performatives with conjunctions and disjunctions

Krifka points out the following contrast:

- (7.36) a. I hereby promise you to sing, and I hereby promise you to dance.
 equivalent to: I hereby promise you to sing and to dance.
 b. #I hereby promise you to sing, or I hereby promise you to dance.

The conjunction of two explicit performatives is again a well-formed explicit performatives (and is self-verifying), while the disjunction of two explicit performatives is odd.

Krifka's explanation for this is complex: He maintains that **and** in (7.36a) is not Boolean conjunction, but a different kind of operator, 'dynamic conjunction', which conjoins speech acts. The infelicity of (7.36b) is explained by the fact that **or** only has a Boolean reading, and that Boolean **or** cannot apply to speech acts.

But it seems we can account for the contrast in (7.36) much more simply. Indeed, the contrast follows directly from the analysis of explicit performatives proposed here.

Conjoined explicit performatives are self-verifying. (7.36a) commits the speaker to the existence of an utterance that commits him to $\text{pep}_{sp}(\textit{sing})$ and, at the same time, commits him to the existence of an utterance that commits him to $\text{pep}_{sp}(\textit{dance})$. But being committed to believe in two such commitments again means having both commitments. But that means that the very utterance of (7.36) can serve as witness to *both* existential claims.

Disjoined explicit performatives are not self-verifying (7.36b), on the other hand, does not commit the speaker to a preference for either singing or dancing. It does commit the speaker to the belief that he is either committed to sing or he is committed to dance. But this doxastic commitment does not entail a commitment

for singing, and it does not entail a commitment for dancing. But then, the utterance in (7.36) cannot serve as a witness for *either* of the disjoined claims.

Krifka's analysis of explicit performatives

Krifka's observations directly follow from the analysis of explicit performatives proposed here. Interestingly, Krifka's analysis of explicit performatives is quite similar.¹²

Recall from Section 6.5) that Krifka assumes that all sentences are headed by illocutionary operators. However, he does *not* take explicit performatives like **I promise to come** to be headed by a promise-operator. Instead, the promise operator occurs within the argument to the usual declarative operator **ASSERT**. That is, for Krifka, the structure of an explicit promise is not (7.37), but (7.38).

(7.37) PROMISE(I come).

(7.38) ASSERT(PROMISE(I come))

The inner promise operator in (7.38) is used descriptively, i.e., (7.38) is an assertion that a promise happens. This already sounds very familiar, especially in light of the fact that Krifka, too, takes assertion to be defined as the taking on of a commitment. He does not spell out formally how the performative effect arises, but he summarizes his conception as follows:

“Now, this proposition states that at this index i [...] the speaker has the promissive obligations to come, and that the speaker has just at this point obtained them. The only way how the speaker can heed the assertive obligation is that the minimal change that created the assertive obligation also created the [promissive] obligation. This is exactly how explicit performatives work: The assertion coincides with the promise.”

¹²Condoravdi and Lauer (2011), which first presented the analysis presented here, was written before Krifka (to appear) was available, and it likewise seems that the latter was written before the former was available. So the two analyses were developed independently, but converged on a very similar conception.

(Krifka to appear)

This is not quite the conception proposed here—in particular, it is not clear how the step “The only way how the speaker can heed the assertive obligation is that the minimal change that created the assertive obligation also created the [promissive] obligation.” compares to the self-verification as it is derived in the present approach.¹³ But it is similar enough to suspect that the account of the interaction of logical operators with explicit performatives made above apply also to Krifka’s conception. If so, then even assuming Krifka’s illocutionary operators does not require assuming illocutionary negation or a speech act conjoining version of **and** to account for explicit performatives.

7.5.3 ‘Illocutionary’ vs. ‘perlocutionary’ verbs

The preceding sections show how, given appropriate assumptions about the lexical meanings involved, sentences with **promise** and **order** can be self-verifying, and how this self-verification fails if the predicates are embedded under negation or disjunction (but not under conjunction), or when they are ‘used reportatively’. But I have not yet addressed the second question that I initially posed: What is the difference between verbs that can be used in this way and verbs that cannot?

We can now answer this question, given the way the self-verification of explicit performatives is explained in the current account. Exactly those predicates can be used in explicit performatives which . . .

- (i) predicate the existence of a communicative event, and
- (ii) the only things they require of this event can be characterized in terms of
 - a. the resulting commitments of the speaker, and
 - b. possibly additional (presuppositional) constraints on the speaker’s attitudes.

¹³This is unclear largely because Krifka leaves it underspecified what it is to have ‘assertive commitments’—he talks about ‘being liable for the truth of *p*’ and ‘the commitment to guarantee that the content of the assertion is true’, but he does not spell out what exactly this means

It is hence clear why, on the present account, verbs like **annoy** and **insult** cannot be used in explicit performatives. Their meanings do not only involve commitments made by the speaker, but reactions of the audience. For (7.39) to be true, John has to have done something (possibly an utterance), and further, John's action must have provoked a feeling of annoyance in Mary.

(7.39) John annoyed Mary.

But then, (7.40) cannot be self-verifying (at least in the way explicit performatives are self-verifying), for the mere fact that (7.40) was uttered does not guarantee that the addressee is annoyed.

(7.40) I (hereby) annoy you.

Similar things can be said about **insult**, which arguably entails that the addressee of the insult takes offense.¹⁴

If we want to, we can now define two classes, the 'illocutionary verbs' and 'perlocutionary verbs'. Illocutionary verbs are verbs of saying whose truth-conditional content only involves speaker-commitments; perlocutionary ones are those whose

¹⁴Actually, that is not quite true. There are uses of **insult** which deny an effect on the hearer (I am grateful to Paul Kiparsky for supplying this example):

(i) He insulted the audience in Urdu, but they had no idea what he was saying.

Somewhat mysteriously, such uses appear to require that there is a *third party* which regards the communicative act as offensive for the the addressee. In any case, these uses are generally *not* licensed in virtue of attitudes of the speaker alone, which goes some way to explain the non-performativity of **insult**. If not knowing general relativity is not taken to be a sign of a negative quality by anyone but John, one cannot say (ii), but has to say (iii):

(ii) John insulted them by explaining general relativity(, which only specialists understand anyway).

(iii) John tried to insult them by explaining general relativity(, which only specialists understand anyway).

This contrast suggests that there is something besides the speaker's intention to offend that is necessary for an action to count as an insult.

truth-conditional content involves further effects on the audience.¹⁵ And then, of course, we can go one step further and call acts named by illocutionary verbs the ‘illocutionary acts’ and verbs named by perlocutionary verbs the ‘perlocutionary acts’. The distinction between these two classes of acts, however, still would not figure in any substantive way in our theory of language use, over and above the fact that illocutionary acts are those that can be performed by explicit performatives, while perlocutionary acts cannot.¹⁶

7.6 Comparison to existing approaches

Searle (1989) opens with the observation that “[t]he notion of a[n explicit] performative is one that philosophers and linguists are so comfortable with that one gets the impression that som[e]body must have a satisfactory theory” (p. 535). As might be expected, he then goes on to explain why he thinks existing theories are not satisfactory. In doing so, he mounts a challenge for theories of explicit performatives, esp. those of the assertoric type. To my knowledge, this challenge has not been answered by anyone subscribing to a Searlean conception of speech acts (Although Bach and Harnish (1992) reply to the challenge, their conception of both speech acts and explicit performatives is quite different from Searle’s, as discussed below). I hence confine myself here to briefly summarizing Searle’s argument and his way of dealing with the problem it raises, and to a comparison with Bach & Harnish’s approach, mentioning a few variant approaches along the way.

¹⁵Actually, not all verbs that predicate ‘perlocutionary effects’ are verbs of saying, or even verbs of communication—**annoy** is not, for example, and **insult** may well not be.

¹⁶The class of illocutionary acts would not correspond to the class of illocutionary acts according to Searle or Bach and Harnish. These authors make clear that the existence of an illocutionary verb is not necessary for something to count as an illocutionary act—but perhaps we could approximate their conception by calling illocutionary acts all those acts that *could* be named by a hypothetical illocutionary verb in the sense defined above. But that would not make the class of illocutionary acts any more interesting from the present perspective.

7.6.1 Searle (1989): Explicit performatives as declarations

Searle outlines a set of desiderata for a theory of explicit performatives. Central for our present concerns are the following:

- (a) performative utterances are performances of the act named by the performative verb;
- (b) performative utterances are self-guaranteeing;
- (c) performative utterances achieve (a) and (b) in virtue of their literal meaning, which, in turn, ought to be based on a uniform lexical meaning of the verb across performative and reportative uses.

It should be obvious that the account proposed in this chapter meets all three of these requirements. Yet Searle claims that an assertoric account *cannot* meet all three desiderata. Why?

According Searle, making a promise requires that the promiser *intend* to do so, and similarly for other performative verbs (the *sincerity condition*). It follows that no assertoric account can meet (a-c): An assertion cannot ensure that the speaker truly has the necessary intention. As Searle puts it:

“Such an assertion does indeed commit the speaker to the existence of the intention, but the commitment to having the intention doesn’t guarantee the actual presence of the intention.”

(Searle 1989, p. 546)

This is obviously true and I think Searle’s argument is a death-blow for assertoric accounts of explicit performatives *that take a Searlean view of speech acts*. But I think we should not take such a view. What Searle’s argument shows, to my mind, is that the conditions for promising, etc. should not depend on things like private intentions of the speaker. Instead, what counts for such acts are *public facts*, such as whether the speaker is, or is not, publicly committed to a belief or preference.

The account of explicit performatives presented here takes promising to be constituted by the creation of a public commitment, nothing more and nothing less. And this is why self-verification obtains: If a speaker says, without any indication of insincerity, **I promise p** , he has thereby promised, and his private intentions do not matter one bit. He cannot later claim that when he made his utterance, he did not promise because he failed to have the right intentions. His utterance committed him. Whether he desired or intended to be so committed by his utterance is immaterial.

Searle, however, thinks that intentions are necessary. Instead, he advocates giving up the idea the explicit performatives are assertions, and proposes to view them as *declarations*, similar to utterances such as **The meetings is adjourned**. Such declarations work because there is an extra-linguistic institution that ensures that if the right speaker says that the meeting is adjourned, it thereby is.

Promises, orders, and the like are different from these cases only in that the fact that is created is a *linguistic* fact (namely that an order, promise, etc. happened). Hence, they do not depend on an *extra-linguistic* institution, but only on the institution of language itself.

So far so good, but what kinds of linguistic conventions make it the case that I can declare with **I promise p** , but not with **I insult you**? And further, what makes it the case that **I promise p** can *only* be a declaration (for recall, Searle wants to explain, as I do, why explicit performatives are self-verifying—they cannot be mistaken, they cannot be lies)? This is where things get a bit murky. Searle says we need three ingredients:

“First, we need to recognize that there is a class of actions where the manifestation of the intention to perform the action, in an appropriate context, is sufficient for the performance of the action.

Second, we need to recognize the existence of a class of verbs which contain the notion of intention as part of their meaning. To say that a person performed the act named by the verb implies that he or she did it intentionally, [. . .]

Third, we need to recognize the existence of a class of literal utterances which are self referential in a special way, they are not only *about* themselves, but they also operate on themselves. They are both *self-referential* and *executive*.”

(Searle 1989, p. 551)

The first two of these are reasonably clear. It is not exactly obvious what it is to ‘manifest’ an intention, but I think we can understand this notion as something that (i) one can only do if one has the intention in question and (ii) displays this intention, i.e., makes it clear that one has it.¹⁷ The second element seems unproblematic in general. The third, I must confess, I do not understand. While it is quite clear what Searle means by ‘self-referential’, I am at a loss in understanding what kind of property ‘executive’ is supposed to be, and why explicit performatives have this property, but other sentences do not.

Granting Searle’s conception of speech acts requiring an intention, we can also grant that an explicit performative makes itself true if it ‘manifests the intention’ to perform the act in question. But then we must show that explicit performatives always and automatically manifest the intention to perform the act that they describe. It must be executiveness that ensures this, but how it does so, and where it comes from, is patently unclear. Searle again:

“[A] way to manifest the intention to perform an illocutionary act is to utter a performative sentence. Such sentences are self-referential and their meaning encodes the intention to perform the act named in the sentence by the utterance of that very sentence. Such a sentence is ‘I hereby order you to leave.’ And an utterance of such a sentence functions as a performative, and hence as a declaration because (a) the verb ‘order’ is an intentional verb, (b) ordering is something you can

¹⁷Perhaps the manifestation has also to be caused by the intention in question. In this case, it would be like the conception of ‘expressing’ employed in the accounts of assertion due to Williams (2002) and Owens (2006) briefly discussed in Chapter 4.

do by manifesting the intention to do it, and (c) the utterance is both self-referential and executive[.]”

(Searle 1989, p. 552)

He goes on to say that the executiveness is ‘indicated’ by **hereby**, by which he appears to mean that **hereby** makes explicit the executiveness that would not be explicit in the sentence without **hereby**:

“Performative speaker meaning includes sentence meaning but goes beyond it. In the case of the performative utterance, the intention is that the utterance should constitute the performance of the act named by the verb. The word ‘hereby’ makes this explicit, and with the addition of this word, sentence meaning and performative speaker meaning coincide.”

(Searle 1989, p.552)

We could make sense of this if we recall that Gricean speaker meaning is determined by the speaker’s intentions. So Searle claims that an explicit performative can be ‘executive’ because the speaker intends it to be (which he could make clear by including **hereby**). If so, then the speaker can simply *intend* to make his utterance manifest his intention to promise.

But it is unclear what guarantees that a speaker who sincerely utters **I promise p** has both the intention that his utterance be executive and the intention to promise. Searle appears to think this has to do with the literal meaning of the utterance.

“The literal meaning of the utterance is such that by that very utterance the speaker intends to make it the case that he orders me to leave. [...] Therefore, in making the utterance S manifested an intention to make it the case by that utterance that he ordered me to leave.”

(Searle 1989, p. 553)

Again, this begs the question: Presumably, what makes the literal meaning of the utterance such that by that very utterance the speaker has the right intention is the word **hereby**. But is unclear how the word achieves this. Recall that Searle must ensure that the speaker *cannot* not have this intention. It seems a speaker could say **I hereby promise** *p* without having that intention, perhaps carefully *avoiding* having that intention, so as to not promise what he purports to promise, so that he cannot be later held responsible.¹⁸

Three more quotations. In summing up his argument, Searle says:

“It turns out under investigation that [the question how an explicitly performative utterance can constitute the performance of the act named by the main verb] is the same question as how the literal utterance of these sentences can necessarily manifest the intention to perform those acts.”

(Searle 1989, p. 555)

On Searle’s construal of illocutionary acts, this is indeed the question, but we can clarify it a bit by rephrasing slightly: How does it come about that these sentences necessarily manifest the intention in question? I fail to see how it is answered by:

“You can perform any of these acts by an utterance because the utterance can be the manifestation (and not just a commitment to the existence) of the relevant intention. But you can, furthermore, perform them by a performative utterance because the performative utterance is self-referential to a verb which contains the notion of the intention which is being manifested in that very utterance.”

(Searle 1989, p. 556)

¹⁸One possibility is that Searle thinks that, regardless of whether the utterance actually *is* a promise (i.e., whether the speaker has the right intentions), any utterance of this form commits the speaker—but if so, there is nothing left to explain: Then we just can define promising as taking on the respective commitment, as I have done.

Both these sentences talk about what you *can* do, while the puzzle is why it is that speakers of explicit performatives *must* be performing the act in question. So how do these considerations ensure the following?

“The literal utterance of ‘I hereby order you to leave’ is — in virtue of its literal meaning — a manifestation of the intention to order you to leave.”

(Searle 1989, p. 556)

I conclude that Searle’s account of explicit performatives is unclear. The only way I see it could be right is if we take the intentions he talks about to be the *publicized* intentions of the speaker, which roughly correspond to the public effective preferences in the account presented here and in Condoravdi and Lauer (2011). But once we make this move, Searle’s original argument against assertoric accounts loses its force: If we understand ‘public intention’ as ‘effective preference the speaker is committed to’, self-verification follows directly from the assumption that the declarative explicit performative has the effect that all declarative utterances have. There is nothing left to explain.

7.6.2 Bach and Harnish (1979, 1992)

Bach and Harnish (1979) (B&H79) analyze explicit performatives as ‘indirect speech acts’ on a par with **Can you pass me a salt?** when used in order to get the speaker to pass the salt (rather than elicit information about his abilities). They provide the following Gricean inference pattern that an addressee can (be expected to) go through in order to infer that a speaker’s utterance of **I order you to leave** constitutes an order (the version given here is the one from Bach and Harnish (1992, p. 99)):

- (7.41)
- a. He is saying ‘I order you to leave.’
 - b. He is stating that he is ordering me to leave.
 - c. If his statement is true, then he must be ordering me to leave.

- d. If he is ordering me to leave, it must be his utterance that constitutes the order (what else could it be?).
- e. Presumably, he is speaking the truth.
- f. Therefore, in stating that he is ordering me to leave, he is ordering me to leave.

Bach & Harnish and Searle's challenge

In Condoravdi and Lauer (2011), we lumped B&H79's account with other assertoric accounts that Searle (1989) targeted (just as Searle himself did), which we characterized as accounts that "tr[y] to derive the performative effect by means of an implicature-like inference that the hearer may draw based on the utterance of the explicit performative." (p. 150). And later we say (n. 1): "It should be immediately clear that inference-based accounts cannot meet [Searle's challenge]. If the occurrence of the performative effect depends on the hearer drawing an inference, then such sentences could not be self-verifying, for the hearer may well fail to draw the inference."

As García-Carpintero (2013, n. 14) points out, that this way of criticizing B&H79's account is a bit rash. In Bach and Harnish (1992) (B&H92), the authors stress that they do not claim that, for an order to occur, it is necessary that the hearer draw the inference in question. All that is necessary, they insist, is that the speaker *intend* the hearer to draw this inference. Whether or not the hearer actually does will determine whether or not the order is 'communicatively successful', not whether the order actually occurs. So, on B&H92's account, if a speaker intends to perform an ordering with his utterance of **I order you to leave**, but the addressee fails to draw the inference in (7.41), the speaker has performed an order, but that order has not been communicatively successful.

It should be immediately obvious, however, that this feature of their analysis does not get them out of the problem raised by Searle's challenge: Whether or not an ordering actually happened still depends on the presence of an intention, though the intention for B&H92 is slightly different than the one Searle requires (see below). That means that explicit performatives, on their account, cannot be

self-verifying, as it is possible to utter **I order you to leave** without having the correct illocutionary intention.

B&H92 do not deny this:

“Searle is surely right to insist that for an utterance actually to be a promise requires a further intention, but it is a mistake to think the utterance should *guarantee* the existence of the intention [. . .]”

(Bach and Harnish 1992, p. 103)

It is unclear then, how it comes about that an utterance of **I order you to leave** is sufficient for ordering—and why any such utterance can be described as an order. It seems that B&H92 want to simply deny that it is (“So a performative utterances is no more self-guaranteeing than a speech act of the same type made non-performatively”, p. 104). But then, we should find utterances of explicit performatives that are not performances of the act described by the performative verbs.

B&H92 appear to think there are such cases, given the fact that there are sentences that have the same surface form as explicit performatives that are not performances of the act named:¹⁹

“The trouble is, an unambiguous performative sentence can be used literally to report some habitual act, in which case one is not performing the act named by the verb [. . .]. You can use ‘I promise . . .’ to report the sort of promise you regularly make at a certain time or in a certain situation.”

(Bach and Harnish 1992, p. 98)

¹⁹B&H92 mention another kind of case involving double-channel communication, where one “describe[s] some collateral act, in which case it is not in the utterance that one is performing the act.” As examples of such collateral acts they mention signing one’s name or nodding. But double-channel communication is no argument against a self-verification account like the one defended here: In case of double-channel communication, there happen to be two witness for the existential claim made, the utterance and the collateral act, instead of just one.

But this argument, which is reprised by García-Carpintero (2013), rests on a linguistic confusion: If a sentence of the form **I promise that ...** can be used both as an explicit performative and to report one's promising-habits, the sentence is not 'unambiguous' in the sense that it has the same denotation on both these uses. On the habitual use, the denotation of the sentence involves some kind of generic quantification over times, occasions, cases, situations, or something of the kind, on the performative use, it does not. But then, this is not an argument against the claim that explicit performatives are self-verifying.

In conclusion, Bach and Harnish's account does not derive self-verification, and hence does not explain why an explicitly performative utterance is *sufficient* for the act to be performed. At best, their account explains why an explicit performative *can* be a performance of the act, not why it *must* be.

Bach & Harnish's conception of promising

Despite their failure to meet Searle's challenge, B&H79's (and B&H92's) account of promising and similar speech acts, and hence their account of explicit performatives involving such acts, is quite different from Searle's and the conception of such acts in the present dissertation.

In a nutshell, the difference is this: On B&H79's account, promising does not necessarily create commitments. To see this, consider their definition of promising:

(7.42) **Promising according to B&H79** (p. 50)

In uttering *e*, *S* promises *H* to *A* if *S* expresses:

- i. the belief that his utterance obligates him to *A*,
- ii. the intention to *A*, and
- iii. the intention that *H* believe that *S*'s utterance obligates *S* to *A* and that *S* intends to *A*.

First, consider the simple case in which the speaker *S* in fact has all the attitudes he expresses, and that the addressee *H* believes that he does: In that case, it will be true that the speaker *believes* that he is obligated, but not necessarily that he also *is*

obligated.²⁰

Things get worse once we see that, on B&H92's account, a speaker may express attitudes that he does not have—he only needs to R-intend that the addressee takes his utterance as a reason to think that he has it (cf. the discussion B&H79's conception of expressing in Section 4.3.1). But that means that, under definition (7.42), a speaker can be said to have promised if he doesn't even believe (but R-intends his addressee to think he believes) that he is obligated to *A*.

This is apparently intentional:

“Commissives are acts of undertaking obligations, but to undertake an obligation is not automatically to create one, even if *S* uses a performative like ‘I promise.’”

(Bach and Harnish 1979, p. 125)

I shall not try and sort out here the difference between ‘undertaking’ an obligation and ‘creating’ an obligation, or how B&H92 intend to account for the fact that promises *do* create obligations, at least in many cases. Doerge (2009) contains an attempt to sort through these terminological issues. Instead, I will only note that I concur with Doerge when he says that “I cannot promise something without, indeed, actually creating an obligation to do the promised thing.” But B&H79/B&H92's account of explicit performatives, at best, can explain why an explicit performative counts as a promise in the sense of the definition in (7.42). Their account can explain why an explicit performative *can* (but does not need to) express the attitudes listed in i.–iii. The account cannot explain why such an utterance creates a commitment. But if the creation of a commitment is necessary for the truth of (7.44), as I believe it is, we are left without an explanation why the utterance in (7.43) can be reported using the sentence in (7.44).

(7.43) John: I promise to come.

(7.44) John promised to come.

²⁰Also note that even if *S* has the attitudes in i.–iii., he himself need not think that it is his utterance that creates the obligation. He only needs to intend that the addressee thinks that.

7.7 Conclusion

This chapter has shown how the self-verification of explicit performatives arises straightforwardly from the theory of clause-typing introduced in the previous chapters, given the right lexical meanings for the verbs that feature in them, which can be derived from the study of *reportative* uses of these verbs. The account makes various welcome predictions about Boolean combinations of explicit performatives, and explains what the difference is between verbs that can function in explicit performatives and those that cannot.

Chapter 8

Exclamatives and expressives

I have argued that we should understand the sentential force of declaratives, imperatives and interrogatives in terms of the commitments that utterances of sentences of these types induce, in virtue of normative conventions of use. Does that mean all clause types should be subject to such normative conventions?

I think the answer is ‘no’. English features a clause type that I do not think is conventionally associated with commitment: exclamatives.¹ These come in at least three sub-types: **wh**-exclamatives, such as (8.1a), nominal exclamatives, such as (8.1b) and inversion exclamatives, such as (8.1c).

- (8.1) a. (My,) How high this building is!
b. (My,) The height of this building!
c. (Boy,) Is this building high!

I will argue that utterances of exclamatives do *not* result in speaker-commitments. Instead, I propose that exclamatives are associated with the kind of non-normative conventions that Lewis hypothesized for clause-types more generally (cf. Section 4.2).

That is, I claim that there are (at least) two quite different ways in which a sentences type can be connected to use. Some types are associated with *normative*

¹Much of the material in this chapter is informed by joint work with Anna Chernilovskaya and Cleo Condoravdi on exclamatives (Chernilovskaya, Condoravdi and Lauer 2012, in prep).

conventions: their force consists in the commitments they induce. Others are associated with *Lewis conventions*, their force arises differently. From a certain point of view, this may seem unfortunate. It would be more uniform to assume that all clause types are associated with the same kind of convention of use. But we also want to ‘carve nature at its joints’—we want to model things that are alike in a like fashion, and things that are not, we want to model differently. And I think that exclamatives work quite differently from the commitment-based declaratives, interrogatives and imperatives.

In my view, exclamatives are similar to the non-sentential utterances that Kaplan (1999) discusses, which have been come to be called ‘expressive’. Examples are **Ouch!** and **Oops!** While utterances of these are, on some level, informationally equivalent to the declaratives in (8.2b) and (8.3b) (the ‘paraphrases’ are Kaplan’s), they intuitively convey their implication in a different way:

- (8.2) a. Ouch!
b. I am in pain.

- (8.3) a. Oops!
b. I just witnessed a minor mishap.

In an oft-cited passage, Kaplan articulates his intuition that these items work quite differently from other expressions usually studied by linguistic semanticists:

“When I think about my own understanding of the words and phrases of my native language, I find that in some cases I am inclined to say that I know what they *mean*, and in other cases it seems more natural to say that I know how to *use* them.”

(Kaplan 1999, p. 4)

Another item Kaplan cites is **goodbye**, about which he writes:

“It is odd to even ask what it *means*. [...] What one needs to know about ‘goodbye’ is that it is an expression conventionally used at parting.

Put in terms of the present discussion, what one needs to know about ‘goodbye’ is how it is *used*.”

(Kaplan 1999, p. 4)

Of course, as discussed in the previous chapters, for *any* sentence type, speakers need to know how they are *used*, and that this is linguistic knowledge. But the difference between exclamatives and the other sentence types I have discussed can be articulated, in analogy with what Kaplan says, as follows: While for declaratives, imperatives and interrogatives, speakers know what their normative effects are, for expressives, speakers know *when* they are used.

8.1 Exclamatives as an expressive clause type

8.1.1 The two implications of exclamatives

Focussing on **wh**-exclamatives for concreteness, we can observe that these have two implications, which we may call, following Castroviejo Miró (2008), the *descriptive* and the *expressive* implication:

- (8.4) How high this building is!
- a. *Descriptive implication*: Sp believes that the building is (very) high.
 - b. *Expressive implication*: Sp is surprised/amazed/struck/awed/... by the height of the building.

Chernilovskaya et al. (2012) demonstrate that neither content behaves as if it is asserted in the way the content of declaratives is asserted (an observation that had previously been articulated, but not sufficiently supported), but that just raises the question how either of these contents is conveyed.

Here, I want to focus on (8.4b), the expressive implication, to the exclusion of the descriptive implication.² The expressive implication is a lot like what is conveyed

²Zanuttini and Portner (2003) take the descriptive content to be a semantic presupposition while Rett (2011) treats it as an appropriateness condition in the style of Searle.

by **ouch**: The utterer of (8.4) conveys that he is in a certain mental state, or has just undergone a certain mental event. What makes exclamatives different is that they have compositionally-determined contents. And the attitude expressed by the exclamation is *directed towards* that content. It is about that content.

What kind of content? Various kinds of denotations have been proposed in the literature on exclamatives, such as sets of propositions (Zanuttini and Portner 2003, Gutiérrez-Rexach 1996), propositions (Castroviejo Miró 2008) and properties of degrees (Rett 2008, 2011).

English exclamatives are restricted in a way that exclamatives in other languages are not. They have to obey what Rett (2008) calls the DEGREE RESTRICTION: They must be about a scalar value, i.e., a degree, number or amount. This is so even if the exclamation does not contain overt degree morphology, as in (8.5). In such a case, a scalar dimension has to be fixed contextually.

(8.5) What ideas she came up with in one afternoon!

Sp is surprised/struck/. . . at the number, ingeniousness, stupidity,. . . of the questions she came up with.

Consequently, English **wh**-exclamatives can only be formed using those **wh**-words that can have degree readings—**how** and **what**.³ So regardless of what we take the compositionally-determined denotation of English **wh**-exclamatives to be, we need to assume that this denotation involves degrees—if it is a proposition, it must be a proposition about degrees, if it is a question-denotation, it must be the kind of denotation that degree questions have, and so forth. For concreteness, I will follow Rett (2008, 2011) and assume that exclamatives denote properties of degrees.⁴

Then exclamatives convey that the speaker has a certain attitude, or just underwent a certain mental event, that is directed to the property of degrees. For ease

³Exclamatives in other languages do not have this restriction (Chernilovskaya and Nouwen 2012): Dutch, German, Russian and Modern Greek allow **who**- and **which**-exclamatives, which get non-scalar readings.

⁴Much like discussed for imperatives and interrogatives in Section 6.4, the conventions of use I will propose here could be reformulated to accommodate different denotation types. It is noteworthy that Rett (2011, p. 431) herself derives a proposition from the degree property on the illocutionary level.

of talking, I will refer to this attitude as ‘surprise’, even though the precise nature of the attitude is much less specified: There are uses in which what is expressed is not surprise (or that the speaker’s expectations got subverted), such as (8.6).

(8.6) [A and B are spending a relaxed Sunday afternoon in B’s garden, enjoying the weather. Suddenly:]

A: What a beautiful garden you have!

Intuitively, what A conveys with his utterance is not that he had not expected the garden to be so beautiful (we can imagine he has known the garden for a long time), but instead that he was suddenly struck by an appreciation for its beauty.⁵ But ‘surprise’ is close enough in many cases, and it is much more convenient to have a single name for what is expressed by an exclamation.

8.1.2 The nature of the expressive implication

So, an exclamation like (8.7) conveys that the speaker is surprised about the (maximal)⁶ degree of height the building possesses. But how does it convey that?

(8.7) How tall this building is!

Rett (2011) employs a Searle-style ‘counts-as’ rule that says that an utterance of an exclamation counts as the expression of surprise. While this sounds correct, it is not clear what it tells us about exclamations with respect to a theory of language use. What does ‘expressing’ amount to?

Commitment to be surprised?

Motivated by a desire for uniformity, we might just think that exclamations *commit* their speakers to being surprised, just as declaratives commit their speakers to

⁵In Chernilovskaya, Condoravdi and Lauer (in prep.), we circumscribe this ‘attitude’ more generally as one that (i) is directed towards a content; (ii) is intersubjectively accessible; and (iii) is constituted by a mental *event*, rather than a mental state.

⁶It is an open question whether the orientation to the maximal degree is semantically encoded, or arises pragmatically.

beliefs and imperatives commit their speakers to preferences. That is, we could hypothesize a convention like (8.8):

(8.8) EXCLAMATIVE CONVENTION (normative version)

When a speaker utters a **wh**-exclamative with content φ , he thereby commits himself to act as though he is surprised by φ .

This may sound sensible—surely, if one is surprised, that may affect the way one acts (one might gasp, for example, or utter an exclamative). But it starts to look a little odd once we spell out what ‘act as though’ means:

(8.9) EXCLAMATIVE CONVENTION (normative version, explicated)

When a speaker utters a **wh**-exclamative with content φ , he thereby commits himself to choose his actions in such a way that is consistent with his being surprised about φ .

In the way I have been talking about action choice, it applies to *intentional* action, and it is about how one picks actions based on one’s preferences and beliefs. It is not quite clear how ‘being surprised’ would figure into such intentional action choice. In itself, this is not a strong argument against (8.8)—perhaps it just indicates that we have to adjust the notion of action choice in some way. But I think there is also something conceptually off about the idea that exclamatives commit their speakers to an attitude such as surprise.

Here is why: Surprise is not a *stable* attitude. If I am surprised today that Ariana Huffington used to be a member of the Republican party, this does not mean that I will be surprised tomorrow—indeed, it is quite likely that I will get used to the thought, and cease to be surprised. And this is generally true about most instances of being surprised. There may be exceptions—a fact so baffling that it never ceases to surprise me—but those exceptions are rare. Belief and (effective) preference are different: While we are aware that our beliefs and preferences can change, most of the time, we expect—at least for a large subclass of our beliefs and preferences—that they will not, at least not in the foreseeable future.

At the same time, a crucial fact about commitments is that they are *stable*. That is part of their *raison d'être*, or at least why they are useful: If I am committed to something at time t , I will be committed to that thing at any later time t' , until the commitment gets explicitly rescinded.

It is no mystery, then, why beliefs and preferences are something an agent might want to commit himself to: Not only are commitments good ways to inform others about one's beliefs and preferences, they also enable one's audience to draw useful inferences about one's (likely) future actions. At the same time, taking on a commitment to a belief or preference that one actually has will usually be harmless, because one just commits oneself to do things that one would want to do anyway. If the beliefs and preferences I commit myself to are ones that I both actually have and expect to persist, then the constraints on my future actions simply constrain my actions to do what I expect to be doing anyway.

Things are different, though, if I commit myself to an attitude I expect *not* to persist—in that case, I commit myself to act in a way that I do *not* plan on anyway. So committing to attitudes that are generally unstable would often be a bad thing to do. That is to say, if exclamatives and other expressives were to commit the speaker to such unstable attitudes, we would expect that they are used rather rarely, if at all.

Commitments to a belief about being surprised?

So saying exclamatives 'express surprise' in the sense that they commit the speaker to act surprised is a non-starter.⁷ The considerations in Section 6.2 suggest a commit-based alternative. Perhaps exclamatives simply commit the speaker to a belief that he is surprised—given that surprise is not a persistent attitude, the fact that the doxastic commitment will persist is harmless: The speaker will remain

⁷Maybe we could retreat, and say that the commitments induced by utterances of exclamatives are short-lived, and self-destruct after a short time? We can take on such short-lived commitments with declaratives, provided we are explicit about it: **I am not convinced p is true, but for the next part of the talk I will suppose it is.** Perhaps exclamatives are simply special in that they always create such temporary commitments. That might work, but it runs the risk of taking all content out of the thesis that exclamatives create commitments. If these commitments are always short-lived, how can we even tell whether such a commitment exists?

committed to the belief that he was surprised at the utterance time, but that will not commit him to the belief that he is surprised at later times.

(8.10) EXCLAMATIVE CONVENTION (commitment-to-belief version)

When a speaker utters a **wh**-exclamative with content ϕ at t , he thereby becomes committed to act as though he believes that he is/was surprised about ϕ at t .

This version does not have the problem that the previous version had. Moreover, it allows us to maintain a certain uniformity in the clause type system and it allows us to explain why it is usually reasonable to infer that the speaker of an exclamative *is* surprised: We only need to assume that he would commit to the belief only if he had it, and that he is well-informed about his own mental state.

A first reason to be uncomfortable with this way of accounting for exclamatives is that they would make exclamatives equivalent to *claims* about ones own mental state. This does not quite accord with intuition—there is an intuitive difference between (8.11) and (8.12):

(8.11) How high this building is!

(8.12) I am surprised that this building is so high.

And we find related differences that indicate that the expressive implication is not claimed in the way the content of exclamatives is claimed (cf. Chernilovskaya et al. 2012). For example, claims can generally be challenged as dishonest, even if they concern the internal state of the speaker, which is not possible with exclamatives:

(8.13) a. I am surprised / hungry / in pain.

b. I don't believe you.

(8.14) a. How high high that building is!

b. #I don't believe you.

More importantly, the commitment-to-belief account predicts that one can exclaim by means of an explicit performative. Note that **exclaim** can be used to describe many, if not all, utterances of exclamatives: An utterance of **How high this building is!** can be reported by saying **He exclaimed about the height of the building.** On the commitment-to-belief analysis of exclamatives, this strongly suggests that (8.15a) can be paraphrased as (8.15b).

- (8.15) a. John exclaimed about the height of the building
 b. John made an utterance that committed him to a belief that he is surprised about the height of the building.

But then, (8.16a) should be self-verifying, and hence it should be generally possible to report it with (8.16b), which does not work.

- (8.16) a. John: I (hereby) exclaim about the height of this building.
 b. John exclaimed about the height of the building.

Relatedly, and I think tellingly, it is hardly ever appropriate to describe the utterance of an exclamative with (8.17a) or (8.17b). If we have reason to doubt that the speaker actually was surprised, it would be much more natural to instead say (8.18):

- (8.17) a. He claimed that he was surprised that . . .
 b. He said he was surprised that . . .

- (8.18) He acted all surprised that . . .

Taking this at face value, we could say the following: Uttering an exclamative is a way of acting surprised (as opposed to a way to commit oneself to being surprised), while uttering a declarative is not just a way to act as if one has a belief, (but also a way to commit oneself to a belief).

Exclamatives as associated with a Lewis convention

If this is the right way to think about exclamatives, then exclamatives are not governed by a *normative convention* of the kind I proposed for declaratives, imperatives and interrogatives, but instead by a *Lewis convention* like the one in (8.19).

- (8.19) EXCLAMATIVE CONVENTION (Lewis version)
 Speakers of the speech community (generally) utter a **wh**-exclamative with content ϕ only if they are surprised (struck/awed/. . .) by ϕ .

That is, a competent speaker of English knows that there is a regularity in behavior in the speech community to the effect of (8.19). To treat exclamatives in our dynamic pragmatics, then, we would assume a general (default) belief that a speaker will utter an exclamative only if he is in fact surprised. With this, we will have for any belief state B_{Ad} , for any exclamative $;\varphi!$:

- (8.20) $B_A[\text{utter}(u, S, \ulcorner ;\varphi! \urcorner)] \models S$ is surprised about φ .

Generalizing the convention

The exclamative convention formulated above only applies to **wh**-exclamatives like (8.21a), but the other types of exclamatives—nominal exclamatives like (8.21b) and inversion exclamatives like (8.21c)—appear to have the same kind of conventional use.

- (8.21) a. How high this building is!
 b. The height of that building!
 c. Is that building high!

Given the morpho-syntactic differences between the three kinds of exclamatives, we might be tempted to simply state two additional conventions with exactly the same content, as in (8.22) and (8.23).

- (8.22) NOMINAL EXCLAMATIVE CONVENTION

Speakers of the speech community (generally) utter a nominal exclamative with content φ only if they are surprised (struck/awed/. . .) by φ .

(8.23) INVERSION EXCLAMATIVE CONVENTION

Speakers of the speech community (generally) utter an inversion exclamative with content φ only if they are surprised (struck/awed/. . .) by φ .

This duplication is rather unattractive. As discussed in Section 6.4, if we find a morpho-syntactically heterogeneous class of sentences that all share the same use, it is preferable to assume that all these sentences share the same denotation type, and let the convention reference this type. If we follow Rett (2011) in assuming all three sub-types denote properties of degrees, we can assume the following convention:

(8.24) DEGREE PROPERTY CONVENTION

Speakers of the speech community (generally) utter an matrix sentence with a property of degrees φ only if they are surprised (struck/awed/. . .) by φ .

For this to work, it must be the case that (a) there are no other expressions that denote properties of degrees and that can occur unembedded (unless these expressions also share the same use) and (b) to the extent that exclamatives can occur embedded, the predicates that embed them have to be compatible with the hypothesized denotation.⁸

⁸Whether or not exclamative clauses occur embedded is an open question—Grimshaw (1979) assumed they do, and that predicates such as **surprise** semantically select for exclamatives:

- (i) a. I am surprised at/by how tall this building is.
- b. I am surprised at/by the height of this building.

But this assumption has been questioned, most forcefully by Rett (2011, Section 4.2). This debate shows that it is by no means always obvious whether a certain clause type embeds, and consequently, whether we can find semantic reasons for preferring a certain denotation type over another.

8.2 More expressive meanings

I have proposed to view exclamatives as associated with a Lewis convention, rather than a normative convention. In this section, I want to briefly mention a few other items that can be viewed in this way—or that at least can be plausibly assumed not to be connected to commitment.

8.2.1 Non-sentential expressions

The first candidate is, of course, Kaplan's expressives like **ouch** and **oops**. For these, it is tempting to directly state a convention that governs their use:

(8.25) **OUCH CONVENTION**

Speakers of the speech community (generally) utter **ouch** only if they are in pain.

(8.26) **OOPS CONVENTION**

Speakers of the speech community (generally) utter **ouch** only if they just witnessed a minor mishap.

We can then capture Kaplan's intuition that, for **ouch**, a competent speaker does not know what it *means*, but only how it is *used*: We simply assume that $\llbracket \text{ouch} \rrbracket$ is undefined. Given that we can state the relevant convention of use without making reference to its meaning or content, it would be superfluous to specify a denotation for **ouch**. All that a competent speaker has to know about **ouch** is that it is governed by the **OUCH CONVENTION**.

8.2.2 Non-at-issue meanings

Throughout most of this dissertation, I have talked as if sentences have only *one* semantic value—roughly, what is usually called the 'at-issue' content. But of course, sentences can have other kinds of contents, such as semantic presuppositions and conventional implicatures (CIs) in the sense of Grice (1975) and Potts (2005).

For each kind of content the question arises *separately* what its conventional effect is. Properly construed, the DECLARATIVE CONVENTION that I have proposed specifies that, when a speaker utters a declarative sentence, he becomes committed to its at-issue content. It leaves unspecified what happens with other kinds of content.

Presuppositions. In the current set-up, it seems natural to suppose that when a speaker utters a sentence with semantic presupposition p , he acts as though he is already committed to p . Of course, not all theories of presupposition represent the presuppositions of a sentence as a separate semantic value—the dynamic satisfaction semantics of the kind in Beaver (2001), for example, has only *one* semantic value which encodes both presuppositional content and at-issue content, viz., a partial update function. In the present set-up, such a function is best viewed as an update function defined on commitment states (rather than belief states)—the DECLARATIVE CONVENTION can then be reconstrued as saying that when a declarative is uttered, the denotation of the uttered sentence gets applied to his doxastic commitments.

Conventional implicatures of supplements. For other kinds of content, we must ask, in every particular case, what their effect is. The kinds of CIs discussed by Potts fall into two broad classes: supplements and expressives. Supplements are things like appositive relative clauses (as in (8.27)) and nominal appositives (as in (8.28)). For these, I agree with Potts that they share their main dynamic effect with main clause assertions—i.e., I think supplements create doxastic speaker commitments.

(8.27) Mary, who you met last night, is a painter.

(8.28) Lance Armstrong, a cancer survivor, won the *Tour de France*.

Integrated expressives. Another class of items that Potts (2005) discusses under the label 'CI' are expressive items which are similar to **ouch** and **oops**, but which are, or appear to be, syntactically integrated into a another clause. Potts (2007)

details a number of characteristics that these items have:

- (8.29)
- a. INDEPENDENCE: Expressive content contributes a dimension of meaning that is separate from the regular descriptive content.
 - b. NONDISPLACEABILITY: Expressives predicate something of the utterance situation.
 - c. PERSPECTIVE DEPENDENCE: Expressive content is evaluated from a particular perspective. In general, the perspective is the speakers, but there can be deviations if conditions are right.
 - d. DESCRIPTIVE INEFFABILITY: Speakers are never fully satisfied when they paraphrase expressive content using descriptive, i.e., nonexpressive, terms.
 - e. IMMEDIACY: Like performatives, expressives achieve their intended act simply by being uttered; they do not offer content so much as inflict it.
 - f. REPEATABILITY: If a speaker repeatedly uses an expressive item, the effect is generally one of strengthening the emotive content, rather than one of redundancy.

Supplements (arguably) also have (8.29a-c) but (8.29d-f) are particular to expressive items.

For some of these, in particular the subclass Potts (2005) calls these ISOLATED CIs, the treatment suggested for **ouch** and **oops** seems entirely appropriate. An example is the expressive use of **fucking**, illustrated in (8.30) and (8.31).

(8.30) Thats fantastic fucking news! (Potts 2005, p. 65, (3.40b))
 ~> Sp is in a heightened emotional state.

(8.31) [status message on a social networking site]
 dear america. get a fucking handle on gun laws ASAP.
 ~> Sp is in a heightened emotional state.

For **fucking**, whose contribution does not interact with the content of its host clause in any way, we can simply assume that it is semantically vacuous (i.e., $\llbracket \text{fucking} \rrbracket$ is the identity function) and assume that **fucking** is directly associated with its own convention of use:

(8.32) **FUCKING CONVENTION**

Speakers of the speech community (generally) utter **fucking** only if they they are in a heightened emotional state.

But not all integrated expressives can be treated in such a simple manner (not all expressives are ‘isolated’ in Potts’ sense). Take epithets like **bastard**, as in Kaplan (1999)’s (8.33).

- (8.33) That bastard Kaplan got promoted to tenure.
 \leadsto Sp dislikes Kaplan.

Bastard is different from **fucking** in that it does not just signal an attitude of the speaker, but it signals an attitude *towards* something, and this something is determined by the expression **bastard** syntactically combines with: (8.33) does not signal that the speaker does not like *someone*, it signals that the speaker dislikes *Kaplan*.

What’s more, $\llbracket \text{Kaplan} \rrbracket$ plays two roles in the interpretation of (8.33): It serves as an argument *both* for the expressive contribution of **bastard**, but also for the main clause predicate **got promoted to tenure**. An account of such non-isolated expressive items requires a limited interaction between the specification of the use and the system of semantic composition. Potts (2007) offers a dynamic treatment of such items that can be viewed as consistent with the way I have been talking about expressive items here. The expressive denotations directly manipulate a feature of the context, the *expressive index*, which keeps track of evaluative attitudes that speakers have expressed. Potts stresses this feature of his account:

“Conceptually, it is important to keep in mind that [the expressive denotations] are quite fundamentally different from the usual denotations

in semantics, in that they have access to the context parameter, which normal denotations cannot manipulate. In this specific sense, they are metalingual. I regard this move as necessary to doing justice to the immediacy property.”

(Potts 2007, p. 188)

If we give Potts’ expressive indexes an epistemic construal (i.e., we take them to represent what interlocutors believe about each other’s attitudes), we can see his dynamic treatment as analogous to the Krifka (to appear)-style variant of the current proposal as sketched in Section 6.5. The semantic value is a context change potential that directly updates the epistemic state of the interlocutors.

8.2.3 Expressive vs. prescriptive conventions

I mentioned already that Kaplan (1999) takes expressions like **goodbye** to be of the same kind as **ouch** and **oops**. They certainly are similar in that there is an intuition that a competent speaker needs to know about them how and when they are *used*, rather than knowing what they *mean*.

Goodbye is different from things like **ouch** and **fucking**, however, in that intuitively, it does not express any kind of attitude. That does not mean that we cannot treat it by means of a Lewis convention (such as that in (8.34)), but we could also treat it in terms of prescriptive rules such as those in (8.35) and (8.36).

(8.34) **GOODBYE CONVENTION** (Lewis version)

Speakers in the community utter **goodbye** only when parting.

(8.35) **GOODBYE CONVENTION** (Normative precondition)

One must not: Utter **goodbye** except when parting.

(8.36) **GOODBYE CONVENTION** (Prescriptive rule)

One should: Utter **goodbye** (or . . .) when parting, and not otherwise.

It is not quite clear how to decide which one is more appropriate. (8.36) seems attractive because it might allow us to capture the intuition that if an agent departs

without acknowledging this by means of an utterance like **goodbye**, he will often be felt as ‘having done something wrong’—and this is kind of essential to the meaning/use of **goodbye**. The very function of the expression is to satisfy this requirement.

The same is true for things like honorific markers in languages like Japanese or the formal/familiar versions of pronouns in Romance and Germanic languages. If we analyze these in terms of prescriptive rules, rather than commitments or Lewis conventions, we directly capture the sense that a speaker who misuses them has committed a *faux pas*.⁹

If these considerations are on the right track, there are *three* ways an expression can have a conventionally-specified use: (i) in virtue of a normative convention specifying commitments, (ii) in virtue of a Lewis conventions specifying when typically is used and (iii) in virtue of a prescriptive rule specifying when it should and should not be used.

8.2.4 Determining the correct convention type

Every type of matrix sentence must have a conventionally specified use, or so I have argued. In addition, every *type* of content a sentence can have must have a conventionally specified use, and these uses may vary with the type of (non-matrix) expression that contributes the additional content.

And there are at least two, and quite possibly three, different possibilities for the conventional specification of use. In each case where we want to specify the conventional use of an expression (or one of its meanings), we must hence ask ourselves: How is it specified? By means of a Lewis convention? A normative convention inducing a commitment? A prescriptive rule?

It would be advantageous to have a set of diagnostics, at least rough and ready ones, that can generally be applied to answer this question, but I am afraid

⁹But of course, this is not the *only* way to capture this intuition. Potts and Kawahara (2004) and Potts (2005, 2007) offer treatments of these items that propose that they express attitudes, just as **fucking** and **bastard** do. Such treatments can capture the intuition that a misuse of these items is a *faux pas* by hypothesizing that there are social rules that mandate or prohibit the expression of the attitudes involved.

I cannot offer such a set at present. In the discussion in this chapter and the previous ones, I have made arguments for specific cases, but most of these do not generalize readily. A stringent requirement for retraction is a good reason to favor the thesis that a given expression type (or an implication thereof) is governed by a normative convention, rather than a Lewis convention—but for this to be a strong argument, the retraction requirement has to be more stringent than those due to communicative reasons. In the case of declaratives, the phenomenon of loose talk (Section 4.7) provides a nice test case, but for many other types of expressions, it is not clear that we can find similar phenomena.

Potts (2007)'s properties of expressives may be taken as a guide in the other direction: If the effect of an expression is 'immediate' and 'descriptively ineffable' that may serve as an indication that it is associated with a Lewis convention—but do note that the creation of a commitment, in the present theory, is just as 'immediate' as the expression of an attitude. Similarly, immediacy and descriptive ineffability arguably are also properties (at least possibly) of items that are associated with prescriptive rules—Potts argues that honorifics and formal/familiar pronouns have these properties, but his arguments are not incompatible with the intuition I just articulated that they are associated with prescriptive rules.

In some cases, we may appeal to introspection: Someone who uses German **du** where **Sie** would have been appropriate intuitively has committed a social transgression—suggesting that a prescriptive rule is at play. Someone who says **ouch** even though he is not in pain is not felt to have committed such a transgression, though if he consistently does so, we might start to wonder whether he knows how that word is used *properly*. But what about someone who says **goodbye**, before animatedly continuing the conversation?

In some cases, it may well not matter, and, in line with what I said about the denotational type of imperatives in Section 6.4.3, maybe there are cases where there is no fact of the matter whether a given item is governed by a Lewis convention or by a normative rule—some speakers take it to be one, others take it to be the other, but due to the nature and function of the item, it simply never makes a difference. In other cases, there will be difference in how the item is used, and so it will be

relevant to correctly identify how the use condition in question is conventionally determined. Whether or not this is necessary will depend, of course, on our explanatory goals in a particular case.

8.3 Conclusion

Much of this chapter has been quite tentative. It was aimed at establishing three points:

- (i) There is good reason to think that there are clause types and other expressions that are unlike declaratives, imperatives and interrogatives, in that they do not, by linguistic convention, induce commitments for the speaker.
- (ii) We hence need to ask, for any particular semantic implication that an expression has, how this implication is related to use.
- (iii) While I have proposed that there are two distinct ways this can happen, commitment-inducing normative conventions and Lewis conventions, there may well be other types, such as prescriptive rules.

And even though I have pointed out that the question how the conventional effect is determined may be unanswerable in particular cases, a theory of language use will require an answer in many others. It is not an accident that declaratives and imperatives are often used in planning joint action and to exchange factual information, while exclamatives are usually used to emote, and honorifics and formal/familiar pronouns are used to negotiate social relationships. I believe there is much systematicity to be uncovered once we know which distinctions to pay attention to, and what the space of theoretical possibilities is.

Chapter 9

Conversational implicatures

The present chapter is concerned with CONVERSATIONAL IMPLICATURES, the topic that Grice (1975) was mainly concerned with when he laid the foundations of the approach to pragmatics that has come to be called 'Gricean'. The previous chapters have applied a broadly Gricean perspective to a range of issues concerning the conventional constraints on force that are conditioned by the type of a clause. These chapters illustrate one of the main points this dissertation aims to make, namely that such a perspective is fruitful *beyond* the study of conversational implicatures.

At the same time, an understanding of the conventional force of sentences is a *prerequisite* for studying the kinds of pragmatic inferences that are the concern of Grice's theory of implicature. Grice's theory *starts* from the idea that a speaker uttered a (declarative) sentence in order to convey information, that is, in order to make the hearer believe that the truth-conditional content of the sentence is true. It hence makes sense that we have fixed a theory of clause-typing before we delve into the study of implicature.

The present chapter has two aims: On the one hand, it shows how 'optimization-based' theories of implicatures fit into the framework of dynamic pragmatics. On the other hand, it shows that such optimization-based theories make surprising, and correct, predictions about the nature of conversational implicatures that are at odds with what (Neo-)Griceans have long believed.

By ‘optimization-based’ analyses of implicatures, I mean any analysis that construes conversational implicatures as inferences the addressee can draw by assuming that the speaker chooses his utterance so as to best satisfy a set of constraints—be they a set of conversational maxims, a more over-arching cooperative principle, preferences of the speaker, or a combination of these (Grice’s theory was optimization-based in this sense, and I submit that every theory that deserves the name ‘Gricean’ must be). The last decade in particular has seen a variety of proposals about how this optimization can be modeled formally, drawing on ideas in optimality-, decision- and game-theory (Blutner 2000, Parikh 2001, Blutner 2002, van Rooij 2004, Benz and van Rooij 2007, Jäger 2007, Franke 2009, Jäger and Ebert 2009, Franke 2011, Jäger 2012, a. o.). I do not want to discuss the differences of these approaches, nor do I wish to present an alternative to them, or exhaustively discuss how they fit into the existent literature on conversational implicatures (Franke (2009), in particular Ch. 1, does a good job of concisely situating these formal analyses within the landscape of Neo-Gricean and Post-Gricean approaches to implicatures, and Jäger (2012) provides a useful summary of the further development of game-theoretical models since). Instead, I want to show how the conception of these optimization-based models fits naturally into the present framework.¹ To this end, I offer some preliminary considerations in Section 9.1 and then show in Section 9.2 how some classic cases of implicatures can be modeled in the present set-up using the simplified conception of the decision procedure *Opt* developed in Chapter 5.

¹Though the underlying conception fits naturally, I do not want to claim that integrating any of these analyses into the present framework is trivial: All of them make quite particular assumptions about the representation of beliefs and preferences that the current model does not make.

The decision- and game-theoretic models, in particular, assume a probabilistic representation of graded belief, combined with a representation of preferences by means of real-valued utility functions. It is worth noting that until very recently, the probabilistic aspect of these representations has not been exploited in any essential way. This may make us wonder whether the probabilistic apparatus is necessary to account at least for standard cases of implicatures.

In the context of implicature-like inferences in reference resolution, Degen and Franke (2012), however, have recently employed a variant of Franke (2009, 2011)’s ‘iterated best response’ model to derive and test probabilistic predictions, and Frank and Goodman (2012) have proposed a closely related model that does the same. These results indicate that the added fine-grained representational structure that such models assume may indeed be necessary.

In Section 9.3, I will show that the conception of pragmatic reasoning that underlies the framework of dynamic pragmatics—and indeed, any optimization-based analysis of implicatures, once construed appropriately—predicts the existence of conversational implicature with quite non-standard properties. In particular, it predicts that there are implicatures that are neither *optional* nor *cancelable*, which yet are entirely Gricean in nature. This has important consequences, as these properties have often been used as more or less definitive tests to determine whether an observed implication of a sentence is an implicature or not.

A note on terminology: I use the term ‘conversational implicature’ (which I will abbreviate to ‘implicature’ throughout, as conventional implicatures do not play a role in what is to follow) in an utterance-centered way: An utterance implicates p if it gives rise to the pragmatic inference that p , or if it at least licenses that inference, in its context of use. This use of the term ‘implicate’ is at variance with that of Bach (2006), who insists that implicatures are not inferences. Bach’s conception is speaker-centered, in that it is speakers, not utterances that implicate things, and in that a speaker only can be said to implicate something if he intends to convey it. Though Bach claims my use is based on a misconception, I do not think we actually differ on any substantive issue. It is likely that Bach agrees that what I call ‘implicature’ works the way I say it does—he simply objects to my calling the inferences in question ‘implicatures’. Bach acknowledges that my use of the term ‘implicature’ is common² and it is useful for my purposes, so I see no reason to defer to Bach’s usage.³ In Section 9.2.5, I will briefly return to the issue whether there is a crucial difference, in the present perspective, between implications of utterances that are intended by their speakers and those that are not.

²To pick a random example from a well-known textbook on pragmatics: “For implicatures are not semantic inferences, but rather inferences based on both the content of what has been said and some specific assumptions about the co-operative nature of ordinary verbal interaction.” (Levinson 1983, p. 104)

³The terminological difference boils down to this: An implicature-according-Bach is an implicature-according-Lauer that the speaker intends to convey.

9.1 Preliminaries

This section provides some necessary preliminary considerations for accounting for implicatures in the present system. Section 9.1.1 discusses the role (or lack thereof) that *maxims of conversation* play in an analysis of implicatures as it is conceived of here. Section 9.1.2 discusses the role of alternative utterances in implicature derivations. Section 9.1.3, finally, introduces a second kind of preference (in addition to the *outcome preferences* our system can model already), and introduces the way I will use optimality-theoretic tableaux to display the outcome of the decision procedure Opt in the rest of this chapter.

9.1.1 Maxims and preferences

Grice (1975) derived implicatures on the basis of his *cooperative principle*.

- (9.1) COOPERATIVE PRINCIPLE (Grice 1975, p. 45)
 Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

More specifically, Grice proposed a set of MAXIMS OF CONVERSATIONS, “the following of which will, in general, yield results in accordance with the Cooperative Principle” (p. 45–46):

- (9.2) a. MAXIM OF QUALITY
 Try to make your contribution one that is true.
 (i) Do not say what you believe to be false.
 (ii) Do not say that for which you lack adequate evidence.
- b. MAXIM OF QUANTITY
 (i) Make your contribution as informative as is required (for the current purposes of the exchange).
 (ii) Do not make your contribution more informative than is required.

- c. MAXIM OF RELATION
Be relevant.
- d. MAXIM OF MANNER
Be perspicuous.
 - (i) Avoid obscurity of expression.
 - (ii) Avoid ambiguity.
 - (iii) Be brief (avoid unnecessary prolixity).
 - (iv) Be orderly.

The idea was, roughly, that interlocutors take each other to follow these maxims as best as they can—i.e., that they obey the maxims in choosing which utterances to make. Implicatures then arise as assumptions that are necessary to justify the speaker's utterance in light of the maxims, or, where the maxims are in conflict or ostentatiously violated, in light of the more general cooperative principle.

While subsequent authors, in particular in the 'Neo-Gricean' tradition, often were concerned with revising, extending and clarifying these maxims (early examples are Atlas and Levinson (1981) and Horn (1984)), recent elaborations of Grice's theory, in particular those employing variants of decision- and game-theory, have largely dispensed with direct appeals to Maxims of Conversation and the Cooperative Principle. Instead, a speaker's utterance choice is simply understood, just as any action choice, as the result of the speaker attempting to realize his preferences, based on his beliefs. The preferences in question may, in a particular case, be motivated by general principles of cooperative behavior, but they also may arise from other sources.

This is the view that I will adopt here. In general, I think it is more useful to think of the Gricean maxims as speaker preferences which just happen to be very common, either because they follow from basic assumptions of cooperativity, or because they are due to our make-up as social beings.⁴ Implicature derivation should then be done on the basis of the assumption (appropriate in some and perhaps most contexts, but not others) that these preferences are in place, instead of

⁴This is arguably quite close to Grice's view: Recall that he characterized the maxims as a principles that, if followed, will yield behavior that accords with the cooperative principle.

directly appealing to the maxims. This has two advantages: (i) Not every preference that figures in implicature calculations must be justified as a general principle of human interaction and (ii) our models extend more readily to situations in which there is a significant conflict of interest—i.e., where agents may have (selfish) preferences that defeat some principles of cooperative behavior.

9.1.2 Alternative utterances

Up to now, I have been silent how the set of alternative actions available to an agent at a given time is determined. I simply have assumed that this is determined by the structure of our models (in particular, by the set of available alternative branches of the branching-time model). I will continue to do so in the present chapter, but seeing as implicature calculation depends quite heavily on the reasoning about particular alternative utterances, I want to briefly address the question of how we should think about alternative utterance actions in the present context.

Preconditions on actions

In a model like our current one, it is tempting to think about the set of action-alternatives among which an agent decides as all those actions that are ‘executable’ by the agent at the time. In general, actions come with preconditions. If I am currently in my office, for example, I simply cannot perform the action of walking from my house to the local coffee shop, because the preconditions for this action are not satisfied. So it is tempting to think of the set of actions an agent decides between in a world w at a time t as just the set of all actions whose preconditions are met at t in w .

On some level, this may be correct, but in the case of utterances, it would leave the set of alternative actions extremely unconstrained: In most situations, an agent can, in principle, say just about anything. So it is more attractive to think of the set of alternative utterances the agent chooses between as a subset of those utterances he could, in principle, make. The agent does not decide between *all* alternative utterances, but only those that are salient in the given context. Which actions these

are will depend, in part, on what has happened in the current discourse so far and what the current topic of conversation is. If the agent has just been asked whether he mopped the floor, for example, the utterances **I mopped the floor** and **I did not mop the floor** (or simply **yes** and **no**) can naturally be assumed to be salient, but they need not be the only ones.

The importance of alternative utterances

In the following, I will assume that the set of salient alternative utterances is extrinsically given, but I want to stress that this assumption is not at all innocent. In particular cases (and depending on the choice of the decision procedure Opt and the speaker preferences that can be assumed), it can make a significant difference which alternative utterances are considered. Take for example the classic example of a scalar implicature in (9.3).

- (9.3) Some students came to the party.
 \leadsto Not all students came to the party.

We can summarize the way this implicature is derived classically as follows (this is basically what Franke (2009) calls ‘naive scalar reasoning’):

- (9.4) The speaker uttered (9.3). Alternatively, he could have uttered **All students came to the party**, which would have been relevant information, and is informationally stronger than (entails) (9.3). Hence I can conclude that the speaker does not take this stronger alternative to be true.

But it has long been recognized (at least since Atlas and Levinson (1981)) that this way of accounting for the implicature in (9.3) does not work in quite this fashion if we assume that **Some but not all students came to the party** is one of the alternatives that the addressee takes into account. For if it is, he should analogously reason as in (9.5), deriving the *opposite* of the implicature to be explained.

- (9.5) The speaker uttered (9.3). Alternatively, he could have uttered **Some but not all students came to the party**, which would have been relevant information, and is informationally stronger (entails) (9.3). Hence I can conclude that the speaker does not take this stronger alternative to be true.

In the more recent literature, this problem has come to be called the ‘symmetry problem’,⁵ for obvious reasons. One way to solve it is to simply deny that the **some but not all** alternative is taken into account when reasoning about the speaker’s utterance choice—but this assumption needs to be motivated, seeing as there is no obvious reason to exclude it categorically from consideration.⁶

Alternatives considered by the hearer

Crucially, what matters for implicature calculation is what the *addressee* takes the action alternatives to be (and what the speaker believes the addressee takes the alternatives to be, and so on), not necessarily the alternatives that the speaker *actually* considers. We can probably safely assume that he considers all the possible utterances that he thinks the addressee expects, but he may consider additional ones.

An important wrinkle is that the set of alternative utterances that the addressee considers may well depend on the utterance that is actually performed. For the speaker, it intuitively makes sense that the set of alternative utterances he takes into account is fixed *ex ante*, i.e., before he makes his utterance. But for the addressee, this is not necessarily the case. It may well be that the very utterance he observes makes salient certain alternatives he would otherwise not have taken into account when reasoning about the speaker’s choice of utterance.

⁵The name was coined in unpublished 1997 class notes by Kai von Fintel and Irene Heim.

⁶There are other ways to deal with this problem—e.g., by assuming that there is a ‘manner’ preference favoring the form **some** over the form **some but not all** (which is longer, and more complex), but that there is no such preference between **some** and **all**. Whether or not this solves the problem depends on the way this preference is represented and the precise Opt assumed, and may depend on other factors—e.g., in Franke’s IBR model, if **some but not all** is included in the alternatives, an additional restriction on belief-revision strategies is necessary to derive the implicature, in addition to the assumption that **some but not all** is ‘more costly’ than **some** and **all**. (see Franke (2009, p. 77–89) for discussion).

This may explain why certain implicatures arise fairly reliably—in particular the ones that Grice called ‘generalized conversational implicatures’, such as the **some**→**not all** implicature illustrated in (9.3). The idea is this: Whenever an addressee observes an utterance containing **some**, then, as a matter of cognitive fact, this automatically brings to mind the possibility of uttering the same sentence with **all** in place of **some**, so if the other prerequisites for the implicature are met, the hearer will infer the scalar implicature.⁷

This is a way to capture the idea that these ‘generalized’ implicatures arise because they rely on ‘conventionalized’ or ‘lexicalized’ scales of linguistic expressions (Horn 1972, Gazdar 1979), in a Gricean way: The implicature itself is not conventionalized in any way, but there is a strong association between the members of the scale which ensures that an utterance containing one member of the scale makes the other members of the scale salient.⁸

It is tempting to also assume that this idea of ‘automatic salience of alternatives’ can also help us solve the symmetry problem: Perhaps the *only* alternative expressions that are taken into account are those that are suggested by the utterance that is actually performed. This seems to be what Neo-Griceans like Horn (1972) assumed, and it is also the solution favored by Franke (2009, 2011) in a game-theoretic setting. It is not clear, however, whether this can work as a general solution. It is one thing to suppose that **some** automatically makes salient **all**, but does not make **some but not all** salient.⁹ It is quite another to assume that a certain form is

⁷If this explanation is on the right track, we would expect that we can detect this ‘automatic activation’ of alternative utterances using psycholinguistic methods (e.g., by testing for priming effects). To my knowledge, this has not been done in any systematic manner for the alternatives involved in classic examples of ‘generalized’ implicatures.

⁸Of course, this does not mean that such a lexical association is *required*—in a given context, alternative expressions can also be salient for other reasons, for example, because a certain question was asked. Hirschberg (1985) was the first to stress the importance of contextually-triggered alternative expressions.

⁹Over the years, there have been various proposals for how the set of ‘salient alternative expressions’ is constructed. The classic proposal by Atlas and Levinson (1981) relies on the fact that **some** and **all** are lexicalized, while **some but not all** is not. Horn (1989) requires that the items on a scale (which in turn are used to derive the salient alternative expressions) have the same monotonicity properties, ruling out a scale that contains both **some** and **some but not all**. In recent years, Katzir (2007) and Fox and Katzir (2011) have proposed that the alternatives relevant to implicature calculations are determined with reference to the syntactic structure of the uttered expression, in

generally not considered at all. Considerations of salience can explain why certain forms are always considered whenever a certain other form has been uttered, but it is harder to see how such considerations can generally *exclude* certain alternatives from consideration.

Be that as it may, in the following I will simply assume, as is common practice,¹⁰ that appropriate alternative sets are given and illustrate how, given this assumption, implicatures can be modeled in the system of dynamic pragmatics.

9.1.3 Two types of preferences and a visual representation of Opt

In the rest of this chapter, it will be useful to have a succinct way to represent the interlocutors' preferences and the way they interact in decision making. Even though the definition of Opt I have adopted in Section 5.2.2 (which, again, is adopted here only for the sake of concreteness) is rather simple, it will be useful to have a way to quickly visualize the predictions that are made.

Luckily, there is a familiar visual representation that we can adopt without much ado, given the way Opt is defined: Tableaus as used in Optimality Theory (Prince and Smolensky 1993). For a given w, t such that $\text{Agt}(w, t) = i$, we can let the set of candidates be $\text{Act}(w, t)$ and the set of constraints $\text{EP}_i(w, t)$. Constraint evaluation is done as follows: Candidate action a violates a constraint/preference c iff $B_{i,t,w}[a] \not\# c$.¹¹

particular ruling out from consideration any expression that is syntactically more complex than the uttered expression—just as Atlas and Levinson, Katzir and Fox thus build a preference for simple expressions into the alternative-selection process, while classically, such preferences (or the corresponding MAXIM OF BREVITY) have been taken to play a role reasoning about given alternatives.

¹⁰For example, Sauerland (2004, p. 373): "I should mention that there is one important question that I have nothing to say about here: Namely, the question where quantitative scales come from. I shall here simply take quantitative scales for granted and use them to account for the implicatures of sentences." Similarly, Franke (2009, p. 127): "[t]his issue is strictly speaking orthogonal to the concerns of [game-theoretic pragmatics], which is a theory of reasoning about alternative messages and not a theory of alternatives as such."

¹¹It is noteworthy that this version of OT is *unidirectional* and models things from a *production perspective*, much like work in OT syntax (Bresnan 2000). OT-theoretic work in semantics (Hendriks and de Hoop 2001) instead has tended to take a comprehension perspective, where the candidates are possible interpretations of an observed form. Work in OT pragmatics, instead, has taken a dual perspective with BiOT (Blutner 1998, Blutner 2000).

This takes care of the preferences that we have represented so far as propositions which characterize properties of *outcomes* that the agent prefers. In the context of implicature calculation, however, we will need to be able to represent a second kind of preference, which we may call *action preferences*. These represent preferences between actions the agent has independently from the outcome of the actions. In game-theoretic treatments, such preferences are usually modeled as *costs* that attach to actions. I will use a similar representation here and model action preferences as functions f that map actions to integers, with the interpretation that a is better than a' according to f if $f(a) < f(a')$. The sets $EP_i(w, t)$ thus no longer contain only propositions, but also costs functions. In the OT representations, the number of violation marks an action a incurs on such a constraint will simply be $f(a)$.¹²

Given the simplified language our agents speak, I adopt here a single, simple action preference that models a preference for *shorter* or *less complex* expressions:

(9.6) MINIMIZE

- a. If e is not of the form $\text{utter}(\cdot, \cdot, \cdot)$, then $\text{MINIMIZE}(e) = 0$.
- b. If e is of the form $\text{utter}(\cdot, \cdot, \varphi)$, then $\text{MINIMIZE}(e)$ is the number of symbols in φ .

For more realistic object languages, it is an open question what counts as ‘shorter’ or ‘simpler’: We could count phones, syllables, morphemes, words or maximal syntactic projections, or there could be a more complex, structure-based measure of simplicity or shortness. Which one is adequate is, ultimately, an empirical question. The advantage of our simple, propositional object language is that we can side-step this issue here. I will assume that MINIMIZE is universally present in the speaker’s effective preference structure, i.e., that every speaker strives to minimize the length of his utterances, and that all agents take each other to minimize the length of their utterances. In general, however, this preference will be ranked below the more important outcome preferences. That is, agents strive to minimize the length of their utterances, but they will opt for a shorter form only *if everything else is*

¹²In classical OT terms, these action preferences are thus treated essentially as *markedness constraints*.

equal, i.e., if a shorter form and a longer form are otherwise equally well-suited to achieve the agent's goals. It is here where the lexicographicness of admissible decision procedures (cf. Section 5.2.2) comes into play: This property ensures that lower-ranked preferences are indeed treated as such *ceteris paribus* preferences, and never win out over higher-ranked ones.

9.2 Some classical implicatures

In this section, I will show how some classical implicatures can be derived in the present system. Again, the purpose is not to provide an exhaustive account of these implicatures, but rather to show how optimization-based accounts of implicatures fit into the current system. For concreteness, I use the simple version of *Opt* defined in Section 5.2.2, but this should not be taken as a proposal competing with existing ones. Instead, as before, I want to abstract away from the question what the correct *Opt* is, and instead focus on how, given a suitable *Opt*, we can model implicatures in the present system.

Before looking at some examples of classical implicatures, it is worth noting that we have seen implicature-like reasoning in the previous chapters, even though the inferences in question usually are not discussed as such. One example is the reasoning in Section 6.1.3 for deriving 'advice' uses of imperatives:

(9.7) [Context: Doctor to patient, in a discussion about how to cure patient's current illness.]

Take these pills for a week.

↪ The doctor believes taking the pill is the best way/one of the best ways to cure the patients current illness.

The reasoning presented there is quite similar, in many ways, to classical relevance implicatures like Grice (1975)'s (9.8):

(9.8) [Context: *Ad* is standing next to his obviously immobilized car.]

Ad: I am out of petrol.

Sp: There is a garage round the comer.

\leadsto *Sp* has no reason to believe that the garage is closed, out of petrol to sell, etc.

In both cases, the addressee can reason to a conclusion about the speaker's beliefs, based on contextual assumptions about the speaker's preferences and the utterance that took place. It is a great advantage of having a pragmatic theory that incorporates a suitable theory of clause-typing that these two inferences can be shown to arise in essentially the same way.

Contextual assumptions

I have already introduced the preference MINIMIZE for shorter forms, which I take to be universally present. Throughout, I will also assume the following high-ranked preference, named after the Gricean maxims it approximates, for any $\vdash \varphi \in \mathcal{L}_+$:¹³

(9.9) QUALITY

$$\lambda v \left[v \models \text{pb}_{Sp}(\varphi) \rightarrow \Box_{Sp}\varphi \right]$$

'The speaker is committed to believe φ only if he actually believes φ .'

However, this assumption is crucially different from the assumption that MINIMIZE is universally present: It is an assumption that only makes sense in certain contexts, while I want to think of MINIMIZE as a preference that speakers truly take each other to always have (though it will be frequently defeated by higher-ranked preferences). In previous chapters, we have seen one case where a preference like (9.9) is not present (or at least is outranked by a preference that defeats it): Loose talk.¹⁴ (9.9) may also be absent (or outranked by a contradictory preference) in cases of substantial conflict of interest between the interlocutors, i.e., where the speaker

¹³I once again suppress temporal parameters throughout for the sake of readability. For the same reason, I will often refer to $\text{utter}(Sp, Ad, \varphi)$ as $\text{utter}(\varphi)$, and will write $\text{Opt}(w)$ for the outcome of Opt in w at the time t under discussion, i.e., $\text{Opt}(w) = \text{Opt}(B_{\text{Agt}(w,t)}, w, t, \text{EP}_{\text{Agt}(w,t)}(w, t), \text{Act}(w, t))$.

¹⁴In cases of loose talk, there will generally be other preferences that make the speaker 'stay close to the truth', such as a general, highly-ranked preference for not violating one's commitments (i.e., not being AtFault), together with a possibly lower-ranked preference against retracting existing commitments.

cannot be assumed to be fully cooperative. But here, we focus on cases where there is no looseness involved, and where speaker are taken to be fully cooperative—that is, the cases of cooperative information exchange that Gricean theory has traditionally focused on.

I also assume throughout that the speaker’s pre-utterance belief state satisfies the following:¹⁵

$$(9.10) \quad B_{Sp}[\text{utter}(Sp, Ad, \varphi)] \models \Box_{Ad}\varphi$$

‘If the speaker utters φ , the addressee will come to believe φ .’

9.2.1 A ‘relevance’ implicature

We start with a case that could be viewed as a complete abstraction of a relevance implicature, such as the petrol example in (9.8) above. The level of abstraction will be so high that the modeling is not of much interest in itself, but it serves as useful warm-up to more interesting cases.

I introduce another outcome preference. (9.11) captures the notion that p is ‘relevant’: If p is true, then the speaker wants the addressee to know this.

$$(9.11) \quad \text{INFORM } p$$

$$\lambda v [v \models p \rightarrow \Box_{Ad}p]$$

‘If p is true, then the addressee believes p .’

While we have assumed that the addressee knows that the speaker has the preferences QUALITY and MINIMIZE, we assume that the addressee does not know whether p is ‘relevant’, i.e., whether the speaker has INFORM p as a preference, but that he assumes that the speaker has no other preferences.¹⁶

¹⁵This is actually automatically ensured if the speaker believes that the addressee believes that QUALITY outranks all other preferences of the speaker, as will be the case in the examples we examine in this section.

¹⁶The assumption that the speaker has no other preferences is quite artificial, but it is less problematic to assume that his other preferences are orthogonal to the present utterance choice. As discussed in Section 6.1.3, in full generality, we also want to allow for a notion of *salience* of preferences, so that even possibly-interfering preferences can be ignored when not salient.

This means there are three kinds of worlds in B_{Ad} :

- (9.12) a. Worlds v_{p+rel} such that $\Box_{Sp}p$ is true and
 $EP_{Sp}(v_{p+rel}, t) = \{\text{QUALITY}, \text{INFORM } p, \text{MINIMIZE}\}$
 b. Worlds $v_{p+\neg rel}$ such that $\Box_{Sp}p$ is true and
 $EP_{Sp}(v_{p+\neg rel}, t) = \{\text{QUALITY}, \text{MINIMIZE}\}$
 c. Worlds $v_{\neg p}$ where the speaker does not believe p to be true.

I assume that $\text{Act}_{Sp}(w, t)$ contains four actions:¹⁷

- (9.13) $\text{Act}_{Sp}(w, t) =$
 $\{\perp, \text{utter}(Sp, Ad, p), \text{utter}(Sp, Ad, q), \text{utter}(Sp, Ad, p \wedge q), \text{utter}(Sp, Ad, p \vee q)\}$

The outcome of Opt at v_{p+rel} -type worlds is shown in Fig. 9.1.¹⁸ The only optimal action is $\text{utter}(Sp, p)$. By contrast, the outcome at $v_{p+\neg rel}$ -type worlds is shown in Fig. 9.2. In these worlds, the null utterance \perp is optimal. And clearly, in $v_{\neg p}$ type worlds, uttering p cannot be optimal, given QUALITY . That means we have the following:

- (9.14) a. $\text{Opt}(v_{p+rel}) = \{\text{utter}(p)\}$
 b. $\text{Opt}(v_{p+\neg rel}) = \{\perp\}$
 c. $\text{utter}(p) \notin \text{Opt}(v_{\neg p})$

But then, updating with $\text{utter}(Sp, Ad, p)$ will remove all worlds that are not of type v_{p+rel} and hence:

- (9.15) $\forall w \in B_a[\text{utter}(p)] : \text{INFORM } p \in EP_{Sp}(w)$

¹⁷ \perp is the null action (or any other non-utterance action), $\text{utter}(Sp, Ad, p)$ the one that will be performed, $\text{utter}(Sp, Ad, q)$ another one whose content is known to be irrelevant, $\text{utter}(Sp, Ad, p \wedge q)$ stands for utterances of anything that is longer than p , but entails it, $\text{utter}(Sp, Ad, p \vee q)$ is representative of utterances that are longer than p but do not entail it.

¹⁸With respect to this and the following tableaux, recall that MINIMIZE is an integer-valued *action preference* that is treated like a markedness constraint. In Fig. 9.1, I also mark QUALITY with a ? for candidate q , as it is irrelevant whether or not it is violated (which depends on whether or not the speaker believes q), given that uttering q violates $\text{INFORM } p$.

v_{p+rel}	QUALITY	INFORM p	MINIMIZE
\perp		*	
 utter(p)			*
utter(q)	?	*	*
utter($p \wedge q$)			***
utter($p \vee q$)		*	***

Figure 9.1: Decision in worlds v_{p+rel} where the speaker believes p and takes it to be ‘relevant’

Input	QUALITY	MINIMIZE
 \perp		
utter(p)		*
utter(q)	?	*
utter($p \wedge q$)		***
utter($p \vee q$)		***

Figure 9.2: Decision in worlds $v_{p+\neg rel}$ where the speaker believes p and does not take it to be ‘relevant’

That is, from observing utter(p), the addressee learns that the speaker takes p to be relevant.

9.2.2 A scalar implicature

Scalar implicatures constitute a more interesting case. The example we are going to look at is:

- (9.16) *Ad*: Do you know the current address for John? I need to send him a letter.
Sp: He is in Europe.
 \leadsto *Sp* does not know where in Europe John is.

Let e stand for the proposition that John is in Europe and p for the proposition that John is in Paris (and assume that for all worlds, $p \Rightarrow e$, capturing that p is stronger than e). Given *Ad*’s question (assuming that *Sp* is cooperative), we can assume that throughout *Ad*’s information state, *Sp*’s preferences are given in (9.17):

$v_{\Box p}$	QUALITY	INFORM p	INFORM e	MINIMIZE
\perp		*	*	
 utter(p)				*
utter(e)		*		*

Figure 9.3: Decision in worlds $v_{\Box p}$ where the speaker believes p

$v_{\Box e}$	QUALITY	INFORM p	INFORM e	MINIMIZE
\perp			*	
utter(p)	*			*
 utter(e)				*

Figure 9.4: Decision in worlds $v_{\Box e}$ where the speaker believes e , but not p

(9.17) $\{\text{QUALITY, INFORM } p, \text{INFORM } e, \text{MINIMIZE}\}$

And assume that Sp 's action choices are those in (9.18) (omitting utterances whose content is known to be irrelevant, as those will always be dominated by \perp):

(9.18) $\text{utter}(p), \text{utter}(e), \perp$

Figures Figs. 9.3 to 9.5 show the outcome of the decision procedure at worlds at which Sp knows that John is in Paris, worlds in which he only knows that John is in Europe, and worlds where he does not even know that. The result is in (9.19).

- (9.19) a. $\text{Opt}(v_{\Box p}) = \text{utter}(p)$
 b. $\text{Opt}(v_{\Box e}) = \text{utter}(e)$
 c. $\text{Opt}(v_{\neg\Box e}) = \perp$

$v_{\neg\Box p e}$	QUALITY	INFORM p	INFORM e	MINIMIZE
 \perp				
utter(p)	*			*
utter(e)	*			*

Figure 9.5: Decision in worlds $v_{\neg\Box e}$ where the speaker believes neither e nor p

$v_{\Box p+\neg rel}$	QUALITY	INFORM e	MINIMIZE
\perp		*	
$\text{utter}(p)$			*
$\text{utter}(e)$			*

Figure 9.6: Decision in worlds $v_{\Box p+\neg rel}$ where the speaker believes e and p , on the assumption that p is not relevant

But then, assuming all worlds in B_{Ad} are of one of the three types:

$$(9.20) \quad B_{Ad}[\text{utter}(e)] \models \neg \Box_{Sp} p$$

This is a simple version of a scalar implicature: p is stronger than e , so if the speaker utters e , the addressee can conclude that the speaker does not know whether p .¹⁹

Of course, this inference is *optional*. It does not always arise, but it instead depends on the addressee making the correct contextual assumptions. The crucial assumption here is that the speaker has the preferences **INFORM p** and **INFORM e** —i.e., that p and e are both relevant and that the speaker prefers to communicate the stronger of the two (p) if he believes it to be true. That is, the addressee must assume that the speaker takes the information provided by p , but not by e , to be *relevant* in the context of conversation.

If the addressee's belief state includes worlds in which this is not given, i.e., in which **INFORM p** is not in place, the only decision that changes is that for $v_{\Box p}$ -type worlds (recall that in all those worlds $v_{\Box p+\neg rel} \models \Box_{Sp,t}(p \wedge e)$ since $p \rightarrow e$ is taken to be common geographical knowledge). The decision in those worlds is shown in Fig. 9.6, and we have the following

$$(9.21) \quad \begin{array}{l} \text{a. } \text{Opt}(v_{\Box p+rel}) = \{\text{utter}(p)\} \\ \text{b. } \text{Opt}(v_{\Box e+\neg rel}) = \{\text{utter}(p), \text{utter}(e)\} \\ \text{c. } \text{Opt}(v_{\Box e}) = \text{utter}(e) \\ \text{d. } \text{Opt}(v_{\neg \Box e}) = \perp \end{array}$$

¹⁹Alternative implicatures are possible, if the speaker has (or might have, according to what the addressee knows) other preferences that prevent him from uttering the stronger alternative. Cf. Section 9.3.3 below.

But then, updating with $\text{utter}(e)$ will let both $v_{\Box p + \neg rel}$ and $v_{\Box e}$ worlds survive. So we do not derive the scalar implicature in the addressee's information state:

$$(9.22) \quad B_{Ad}[\text{utter}(e)] \not\models \neg \Box_{Sp}(p)$$

This is of course correct: Scalar implicatures only arise if the stronger alternative is *relevant*:

$$(9.23) \quad [\text{Context: } Ad \text{ and } Sp \text{ are in the US}]$$

Ad: Is John in town?
Sp: He is in Europe.
 $\not\rightarrow$ *Sp* does not know where in Europe John is.

9.2.3 The 'epistemic step'

The scalar implicatures we have derived so far were rather weak:

$$(9.24) \quad \text{Scalar implicature}$$

$$B_{Ad}[\text{utter}(Sp, e)] \models \neg \Box_{Sp} p$$

This 'epistemic' implicature is completely adequate in a case like (9.25).

$$(9.25) \quad \text{John is in Europe}$$

\leadsto *Sp* does not know where in Europe John is.

But in other familiar cases of scalar implicatures, what is perceived is something stronger:

$$(9.26) \quad \text{Some people came to the party.}$$

\leadsto (a) *Sp* does not know that all students came to the party.
 \leadsto (b) *Sp* believes that not all students came to the party.
 \leadsto (c) Not all students came to the party.

But the stronger implications in fact arise only if (9.26) is uttered in the right context: If the speaker is taken to be (honest and) well-informed about which students were at the party, we derive the strongest implication (9.26c). If, instead, the speaker is taken to only *believe* he is well-informed about which students came, we only derive the intermediate one in (9.26b).

Gazdar (1979) derived (9.26b) directly as an implicature (in addition to the weaker (9.26a)), but it seems conceptually more attractive to derive (9.26b) and (9.26c) on the basis of (9.26a), in contexts where this is appropriate, as proposed by Horn (1989). We can derive these additional implications as contextual strengthenings of the epistemic implicature (Sauerland (2004) calls this the ‘epistemic step’). It works much like the inference to the truth of the content in Chapter 2. We characterize the following two contextual conditions:²⁰

- (9.27) a. ‘Speaker has an opinion about p ’
 $\Box_{Sp}(p) \vee \Box_{Sp}(\neg p)$
 b. ‘Speaker is an expert on p ’
 $p \Leftrightarrow \Box_{Sp}(p)$

(9.27a) gives rise to the ‘intermediate’ implicature that the speaker believes the negation of the stronger alternative, while (9.27b) gives rise to the strong implicature that the stronger alternative is false.

Fact 4. Let B_{Ad} as in the previous section, and further $B_{Ad} \models (9.27a)$. Then

$$B_{Ad}[\text{utter}(Sp, e)] \models \Box_{Sp}(\neg p)$$

Fact 5. Let B_{Ad} as in the previous section, and further $B_{Ad} \models (9.27b)$. Then

$$B_{Ad}[\text{utter}(Sp, e)] \models \neg p$$

²⁰Both Sauerland (2004) and Russell (2006) derive the ‘epistemic step’ in essentially the same way as I do here.

9.2.4 Intended implicatures

The implicatures we have derived so far were not (necessarily) *intended* inferences. We did not assume (and did not need to assume), that the speaker preferred to communicate the content of the implicature to the hearer. But of course, implicatures can be intended to be communicated. They can even be the main motivation for an utterance. The system as introduced so far captures this.

Firstly, observe that speakers can be aware of the addressee-inferences their utterances will trigger. The following fact captures this for the strong ‘factual’ implicature from the previous section.

Fact 6. *Let B_{Sp} be such that for all $v \in B_{Sp}$, $B_{Ad,v}$ satisfies the assumptions made in Section 9.2.2 and in addition $B_{Ad,v} \models (9.27b)$. Then*

$$B_{Sp}[\text{utter}(e)] \models \Box_{Ad}(\neg p)$$

Now consider a situation in which the speaker has the preferences the addressee took him to have in Section 9.2.2, but he also has the another one $\text{INFORM } \neg p$, with the obvious interpretation. His preferences are:²¹

$$(9.28) \quad \text{QUALITY} > \text{INFORM } p > \text{INFORM } \neg p > \text{INFORM } e > \text{MINIMIZE}$$

The crucial case to investigate is that of worlds in which the speaker knows $e \wedge \neg p$. Fig. 9.7 shows the tableau on the assumption that the speaker *does not* believe that the addressee will draw the scalar implicature and Fig. 9.8 shows the decision in case he does believe the implicature will be drawn.²²

So if, but only if, Sp believes that Ad 's information state is such that he will draw the implicature, he will prefer uttering e over the longer $e \wedge \neg p$. However, if he takes it to be possible that the addressee will not draw the implicature, he will

²¹The precise ranking of the INFORM -preferences is not essential—indeed, we could assume that all preferences are unranked, with the exception of MINIMIZE , which needs to be subordinate to the outcome preferences.

²²As discussed in Section 9.1.2 above, the simple implementation of implicature-reasoning given here depends on the assumption that (the speaker believes that) the addressee does not take into account the possibility of the utterance $e \wedge \neg p$.

$v_{\Box_i e \wedge \neg p}$	QUALITY	INFORM p	INFORM $\neg p$	INFORM e	MINIMIZE
\perp			*	*	
utter(p)	*		*	*	*
\Rightarrow utter($e \wedge \neg p$)					****
utter(e)			*		*

Figure 9.7: Decision in worlds $v_{\Box_i e \wedge \neg p}$ where the speaker knows that e is true and p is false, and believes the implicature will not be drawn

$v_{\Box_i \neg p \wedge e}$	QUALITY	INFORM p	INFORM $\neg p$	INFORM e	MINIMIZE
\perp			*	*	
utter(p)	*		*	*	*
utter($e \wedge \neg p$)					****
\Rightarrow utter(e)					*

Figure 9.8: Decision in worlds $v_{\Box_i \neg p \wedge e}$ where the speaker knows that e is true and p is false, and believes the implicature will be drawn

instead opt for the longer expression. This arguably accords with intuition.

These are but the first two steps in the kind of back-and-forth reasoning that has played a role in a number of recent accounts of pragmatic phenomena. Franke (2009, 2011)'s ITERATED BEST RESPONSE model and its variants and precursors (e.g. Jäger and Ebert (2009), Degen and Franke (2012)) explicitly define an iterated reasoning chain alternating between the speaker and hearer perspectives to derive implicatures. Recent work by Frank and Goodman (2012) employs a very similar model that makes quantitative predictions which match observed frequencies of implicature calculation in referent-selection tasks, and Vogel, Potts and Jurafsky (2013) use Decentralized Partially Observable Markov Decision Process (Dec-POMDP) models to capture the same kind of iterative reasoning. They apply their model to implement artificial agents that perform implicature-reasoning in the course of solving a collaborative task.

An advantage of the explicitly-stepwise nature of these theories is that they can model *bounded rationality*, i.e., they can allow that whether or not a certain inference is drawn depends on the reasoning capabilities of the interlocutors. This

becomes particularly useful in accounting for some inferences in cases of deceptive language use (as argued by Franke (2008), Franke, de Jager and van Rooij (2012)), as a speaker can be taken to assume that he can ‘outsmart’ the addressee.

The framework developed in this thesis is not in conflict with these approaches, but an abstraction over them. It tries to get by with rather minimal assumptions about the optimization procedure (such as lexicographicness), while the mentioned approaches generally take a rather specific procedure and apply it to a particular phenomenon, perhaps refining the procedure in the process. The focus here is not on the details of the optimization procedure, but rather on the way the belief that such an optimization procedure is followed results in pragmatic inferences on the part of the addressee.

It is noteworthy that the initial scalar inference on the part of the addressee required little pragmatic sophistication (a point that also clearly emerges in the accounts of Franke and Frank and Goodman, where simple scalar inferences like this generally arise on the first iteration step). All that the addressee had to attribute to the speaker was a preference to not say things he does not believe, a preference to inform and a preference to minimize effort.

9.2.5 Unintended implicatures

The scalar implicatures in Section 9.2.2 and the strengthening in Section 9.2.3 were derived without assuming an intention (or effective preference) of the speaker. We did not assume that (the addressee thinks that) the speaker wants to inform the addressee of the truth of the implicature. Only in Section 9.2.4, did we add additional assumptions to this effect. But the basic scalar inference arose without it.

This is empirically adequate, as Horn (2006, p. 37) notes when discussing Ariel (2004)’s example in (9.29): “it must be recognized that in [some] cases we DO implicate what harms, if not defeats, our local purpose.”

- (9.29) The majority decided for peace. Me too.
(bumper sticker, originally Hebrew, spotted 4.2002)

Crucially, someone who reads (9.29) is likely—and intuitively warranted—to infer that not everyone decided for peace. As Ariel points out, in this case, the inference in fact weakens the speaker’s rhetorical point—so we can safely assume that he did not intend to communicate its content. The case is exactly parallel to the scalar implicature we derived before: p is the stronger statement that everyone decided for peace, e is the (asserted) weaker statement that the majority decided for peace. We predict the implicature to arise assuming only the preferences QUALITY, INFORM p , and MINIMIZE. Crucially, what we do not need to assume is a preference INFORM $\neg p$. Indeed, even assuming the opposite preference would not change the outcome of the implicature calculation—as long as QUALITY, the preference for not committing to falsehoods, is ranked above the preference for not informing the addressee of $\neg p$. This, again, accords with intuition. The ‘speaker’ of (9.29) could avoid informing his addressee that not everyone decided for peace—by uttering **Everyone decided for peace**. But he can only do so if he is willing to commit to a belief in the content of this stronger sentence.

From the perspective of a theory of language use, there does not seem to be much of a difference between such scalar inferences that are unintended and those that are intended. If the speaker has the right kind of preference, we may infer, in addition to the scalar inference, that he intended to communicate its content; but that does not impact the way the inference arises in the first place.

Some terminological hardliners (e.g., Bach (2006)) have insisted that a (potential) inference must be intended in order to count as an implicature, at least according to how Grice himself used the term. And while Grice’s characterization of ‘conversational implicature’, quoted below, does not mention the term ‘intention’, it uses ‘implicate’ in the *definiens*. Grice does not give a definition of the simpler term, but the way he introduces it makes clear that he took implicatures to be part of what is *meant* by a speaker. And speaker meaning, for Grice, was what the speaker *intended* to communicate. So perhaps Bach is right that Grice generally thought that implicatures must be intended.

“I am now in a position to characterize the notion of conversational implicature. A man who, by (in, when) saying (or making as if to

say) that p has implicated that q , may be said to have conversationally implicated that q , provided that (1) he is to be presumed to be observing the conversational maxims, or at least the Cooperative Principle; (2) the supposition that he is aware that, or thinks that, q is required in order to make his saying or making as if to say p (or doing so in *those* terms) consistent with this presumption; and (3) the speaker thinks (and would expect the hearer to think that the speaker thinks) that it is within the competence of the hearer to work out, or grasp intuitively, that the supposition mentioned in (2) is required."

(Grice 1975, p. 49–50)

And yet, Grice concludes his 'general pattern for working out of a conversational implicature' as follows:

"he intends me to think, or is at least willing to allow me to think, that q , and so he has implicated that q ."

(Grice 1975, p. 50)

So it seems that Grice agreed that, at least from the addressee's perspective, there is not much of a difference between inferences that the speaker intends and those that he merely tolerates.

In any case, I shall continue to use 'implicature' in the broad sense in which it applies to (potential) pragmatic inferences, regardless of whether the inference is intended by the speaker.

9.3 Need a Reason: Mandatory Gricean implicatures

The previous sections are mainly intended to show how optimization-based analyses of implicatures fit into the current system, by illustrating how a particular optimization procedure (i.e., the Opt-function defined in Section 5.2.2) can be used to derive certain standard implicatures.

In the rest of this chapter, I will argue that this conception of pragmatic inference makes some surprising predictions concerning the properties of implicatures. In particular, I will argue that properly pragmatic inferences can lack certain properties that they are frequently taken to have by necessity. This result is of considerable significance, as these—putatively necessary—properties have often been used as tests to distinguish pragmatic inferences from semantic implications of sentences.

In particular, I will argue that pragmatic inferences can be *mandatory*, in the sense that they arise on every use of a given form. And I will argue that *any* optimization-based account of pragmatic inference, once conceptualized in the right way, predicts the existence of mandatory implicatures. The perspective of the framework of dynamic pragmatics lets us see that this prediction is not at all conceptually suspect. And once we know what to look for, it is not hard to find empirical support for the idea that mandatory implicatures indeed exist.

9.3.1 Optionality and cancelability

In contrast with the claim that I want to establish here—that there are implicatures that are mandatory—it is frequently supposed that all pragmatic inferences are by necessity both *optional* and *cancelable*.

Optionality and cancelability are about the relationship between a linguistic expression and a (potential) implication that utterances of the expression have.²³ For present purposes, the expressions in questions will always be sentences. An implication *i* of a sentence *S* is *optional* if there can be sincere utterances of *S* that do not give rise to *i*; and it is *cancelable* if a sincere utterance of *S* is compatible with a denial of *i* even in contexts in which *i* would normally arise.

These two properties are often grouped under the label *cancelability*, e.g. by Horn (1984):

²³In full generality, it would be perhaps more appropriate to say that optionality and cancelability are properties of the relationship between *utterance types* and their (potential) implications. Generally, though, the relevant utterance type can be usefully identified by the linguistic expression that is uttered.

“[...] the aforementioned inference goes through in unmarked contexts, but it may be cancelled—explicitly [...] or implicitly (by establishing the appropriate context, [...]).”

(Horn 1984, p. 13)

For what follows, it will be convenient to have a separate label for the two properties—and I prefer the term ‘optionality’, as I do not subscribe to Grice’s view that (some) implicatures arise by default unless the the context prevents them—as formulations such as ‘the context cancels the implicature’ suggest.

Many implicatures *are* optional and cancelable, of course. We saw in the last section how standard scalar implicatures depend on the stronger alternative expression being relevant, and hence will not arise in a context in which the information provided by this alternative is not relevant. And they are cancelable—an utterance of (9.30) is compatible with an explicit denial of the scalar inference, as in (9.31):

- (9.30) Some students came to the party.
 \leadsto *Sp* does not know that all students came to the party.
 \leadsto Not all students came to the party.

- (9.31) Some students came to the party. In fact, all of them did.

Hirschberg (1985) pointed out that it is actually strange to speak of the implicature being ‘canceled’ in (9.31), on the Gricean assumption that implicatures are part of speaker meaning, and hence must be intended: Presumably, the speaker of (9.31) did *not* intend to convey that not all students came to the party when he uttered the first sentence. But then, if an implicature is present only if the speaker intended the addressee to recognize it, then *there is no implicature to be canceled* in (9.31).

Of course, there still is a (potential) pragmatic inference that the hearer may have drawn when the speaker uttered the first sentence. And it is this inference that the speaker intends to prevent, or ‘call off’ with his second sentence. So speaking of implicatures being canceled already presupposes a view on which implicatures need not be intended.

But even setting this issue aside, there is a tension in talking about cancelation of pragmatic inferences in optimization-based theories: By assumption, speakers only make utterances that best satisfy their goals and preferences. Now, it seems that in (9.31), we must assume that the speaker wants to avoid making the addressee believe that not all students came. If he would not mind the addressee forming this belief, why did he utter the second sentence? But then, if he chose his previous utterance optimally, should he not have opted for **All students came to the party** in the first place? It seems that optimization-based theories predict that cancelation *never* happens.²⁴

Mayol and Castroviejo (2013) articulate this problem for the traditional Gricean account (which, after all, is an optimization-based account):

“Why do [cancellations] exist at all if, within the Gricean program, they should be viewed as uncooperative? That is, how can we make sense of a discourse where the same speaker first utters a weaker statement and, immediately after, a stronger one? Why did he not utter the stronger statement to begin with?”

(Mayol and Castroviejo 2013, p. 85)

Mayol and Castroviejo (2013) also note that, despite the wide-spread belief (cf. below) that cancelability is a necessary property of conversational implicatures, and the fact that cancelability is often used as a test to determine whether a given implication is an implicature, cancelation “has hardly been studied itself”.

Optimization-based theories can avoid predicting the complete absence of cancellations if they somehow can explain why optimization yields a different outcome for the first and second sentence in (9.31).²⁵ There are multiple ways of doing

²⁴I am grateful for Chris Potts (p.c.) for drawing my attention to this fact.

²⁵There is an alternative route to go: Assume that speakers do not always *make* optimal utterances, even though their audiences believe they do. On this construal, cancelation triggers belief revision on the part of the addressee, making him conclude that the speaker’s first utterance was not optimal after all—this is a possible way to construe such cases, but it makes one wonder why hearers persist in their belief that speakers make optimal utterance choices after they have heard them cancel implicatures multiple times.

so, which all involve the notions of salience, awareness or attention. One way to account for the cancelation in (9.31) is to assume that, when uttering the first sentence, the speaker was unaware that (the addressee might think that) the stronger alternative **All students came to the party** is relevant, but that this occurred to him after uttering the sentence.²⁶ As we saw, the addressee will only infer the implicature if the stronger alternative is supposed to be relevant. So, when the speaker made his first utterance, his choice was optimal, because he did not think that the implicature would arise. If he then becomes aware of the (potential) relevance of the stronger alternative, and hence of the possibility that the addressee drew the implicature, we understand why his second utterance is optimal, too.

Similarly, we could assume that, in choosing his first utterance, the speaker simply did not take into account the possibility that he could also have uttered the stronger alternative, but that this occurred to him after the fact. This is slightly different than the assumption that he considered the stronger alternative, but deemed it irrelevant. In the latter case, he failed to take into account one of his preferences; in the former, he failed to take into account an action-alternative altogether. But the result will be the same: As the stronger alternative did not occur to him when making the first utterance, that utterance was optimal, while after it occurred to him, it was optimal to utter the stronger alternative, canceling the implicature.

If this is on the right track, then we can fully account for cancelation only once we have a representation of salience (or awareness) and its dynamics. But this limitation is no hindrance for our present concerns. For note that in any case, the assumption must be that the speaker's conception of the context shifted between his two utterances: While making his first utterance, the speaker took himself to be in a context in which his utterance would not give rise to the implicature; while making his second utterance, he took himself to be in a context in which the implicature did arise (or in which it is at least possible that it did arise). That means that optionality is a necessary condition for cancelability. An implicature *i*

²⁶This possibility seems to be more or less what Mayol and Castroviejo (2013) have in mind when they articulate their central constraint on implicature cancelation (p. 88): Canceling is only possible if the Question Under Discussion (QUD in the sense of Roberts 1996) that the cancelation addresses is different from the one that the previous utterance addressed.

of a sentence *S* is cancelable only if we can fix contextual assumptions such that an utterance of *S* does not give rise to *i*. As I will argue that there are implicatures that are non-optional, it follows immediately that they are also non-cancelable.

Grice himself thought that being cancelable is a conceptually necessary property of implicatures as he understood them:²⁷

“Finally, we can now show that, conversational implicature being what it is, it must possess certain features: 1. [...] it follows that a [...] conversational implicature can be canceled in a particular case.”

(Grice 1975, p. 57)

And we find little disagreement in the literature. Unsurprisingly, we find statements to the same effect in standard textbooks, e.g., the classic Levinson (1983) and the more recent Kadmon (2001):

“But the implicature, like all implicatures, is defeasible [...]”

(Levinson 1983, p. 138)

“It is uncontroversial that conversational implicatures are *defeasible*.”

(Kadmon 2001, p. 6, emphasis in original)

And we find it echoed in the current literature, where it is used to argue against the pragmatic status of certain inferences (cf. Section 9.3.5):

“Within the Gricean theory, scalar implicatures are pragmatic inferences. Hence, they have a weak status: they are optional, cancellable, and suspendable.”

(Magri 2009, p. 253)

²⁷I elide Grice’s reasoning here for dramatic effect, and will return to it below, in Section 9.3.4. The strong statement in this quote is quite uncharacteristic for the hedge-prone Grice. Later in Grice (1978), he was more cautious (emphasis mine): “I *think* that all conversational implicatures are cancelable.”

So the idea that conversational implicatures are always optional and cancelable was there from the start, it is asserted without hedging in standard textbooks, and it is used, as a critical test, without arguing for it. I think it is safe to say that the idea is wide-spread and generally accepted.

And yet I shall argue that it is mistaken: A Gricean conception of pragmatic inference is quite compatible with such inferences being mandatory.²⁸ In the following sections I will argue that there is a clearly circumscribable class of implicatures that are neither optional nor cancelable, and hence are mandatory. I call these ‘Need a Reason’ (NaR) implicatures, for reasons that will become clear shortly. I in Section 9.3.2, I will discuss a simple example of an implicature of just this kind, and I will show how this implicature comes out as mandatory in the present framework in Section 9.3.3. Subsequently generalize the conditions under which this happens in Section 9.3.4.

9.3.2 The ignorance implicature of disjunction

In the present section, I want to argue that there is an implication, which typically is assumed to be an implicature, which is indeed neither optional nor cancelable: The ‘ignorance’ implication of unembedded disjunction. When a speaker utters a sentence containing an unembedded disjunction, such as (9.32a), his audience is often licensed to conclude that the speaker does not know which of the disjuncts is true.

- (9.32) a. John is in London or he is in Paris.
 b. Speaker does not know that John is in London.
 c. Speaker does not know that John is in Paris.

²⁸Hirschberg (1985, p. 24–32) quite clearly saw that optionality and cancelability do not follow at all from the way Grice defined conversational implicatures, and hence proposed to require that implicatures must be cancelable *by definition*.

This implication is quite robust, at least in cooperative contexts. It is so robust, in fact, that Zimmermann (2000) took it to be an *entailment*.²⁹ But, as I will show, we can explain the robustness of this implication perfectly well in purely Gricean terms, on the assumption that it is a conversational implicature.

If one accepts that the implication is an implicature, it may seem like a garden-variety case of a scalar implicature: There are alternative, stronger sentences (viz., **John is in London** and **John is in Paris**) that the speaker could have uttered instead, and so we conclude that he does not have sufficient evidence for the propositions expressed by those sentences. Isn't this just like a speaker uttering **John is in Europe**, where the audience can infer, due to the existence of an alternative, stronger statement (**John is in London**) that he does not know that the stronger statement is true?

The cases are indeed similar, but the implicature of disjunction is different in that it cannot be straightforwardly called off, at least not in the way this is usually done with scalar implicatures—namely, by asserting the alternative that triggered the implicature:

(9.33) ??John is in London or he is in Paris. In fact, he is in Paris.

Intuitively, (9.33) is odd because if the speaker knows that John is in Paris, we are left to wonder why he uttered the disjunction in the first place. Even though I have called it the 'ignorance' implicature of disjunction, the content of the implicature is not what is given in (9.32), but something more general. That is why it is possible to conjoin (9.32a) with an assertion that the speaker knows which disjunct is true:

(9.34) John is in London or he is in Paris. And I know which one.

Grice (1978) assumes that (9.34) is an instance of implicature cancelation. Instead, I want to view the second sentence as excluding a possible strengthening of a more general implicature. Note that, when (9.34) is uttered, the audience will still need to infer a reason for why the speaker uttered the disjunction rather than just asserting

²⁹Zimmermann was largely concerned with *embedded* uses of disjunctions, but his account predicts that (9.32a) entails both (9.32b) and (9.32c).

one of the disjuncts—a natural one, in this case, is that the speaker is unwilling to share this information. In full generality, then, we can characterize the implicature as follows:

- (9.35) John is in London or he is in Paris.
 ~> The speaker had a reason for not uttering either disjunct.

Why would *or* trigger such an implicature? Eckardt (2007) puts it like this:³⁰

“In using a disjunction, the speaker necessarily has to mention two properties which are usually more specific. These properties are presented as salient and relevant. The simpler sentences are salient alternative utterances in context. The hearer hence will look for a reason why the speaker chose a more complex expression in order to give less information.”

(Eckardt 2007)

In a nutshell, the reasoning is this: There is a general preference for uttering shorter or less complex sentences. This preference can be overridden if the speaker has reasons to utter the longer sentence. In many cases, a speaker who chooses to make a longer utterance (or multiple utterances instead of one) will be motivated by his desire to convey more information. However, when uttering an unembedded disjunction instead of one of its disjuncts this cannot be his motivation, as the disjunction conveys *less* information than the shorter alternatives. So the speaker must have had another reason for not uttering either of the disjuncts.

It is clear, then, why the implicature cannot be canceled by asserting one of the disjuncts: If the speaker is willing and able to assert one of the disjuncts, why did he utter the longer, more complex disjunction in the first place? And we can also see why the implicature is not *optional*: If the preference for a shorter form is always in place (though it may be overridden), then whenever a speaker utters a

³⁰Like Zimmermann (2000), Eckardt (2007) is mainly concerned with embedded uses of disjunctions, but her reasoning applies also to the unembedded uses discussed here.

disjunction, he must have had a reason to avoid the otherwise-preferred shorter disjuncts.

It is useful to consider circumstances in which the disjuncts themselves are not more relevant than the disjunction, i.e., contexts in which the information which disjunct is true does not need to be conveyed. Suppose that all that matters, in a given context, is whether John is in the USA so that Mary can call him. Suppose further that Sue knows that John is in London. Intuitively, we would expect Sue to utter (9.36a) and not (9.36b):

- (9.36) a. John is in London.
 b. John is in London or Paris.

This is so despite the fact that the information that (9.36b) provides is sufficient, given Mary's informational needs. If Sue were to utter (9.36b), Mary would likely and appropriately infer that Sue does not know where John is (or that she has a reason not to share this information), *even though*, by assumption, this information is not relevant.

The case is quite different with a run-off-the-mill scalar implicature. Suppose, in the same context, Sue instead utters (9.37).

- (9.37) John is in Europe.

Even though (9.37), in an appropriate context, would implicate that Sue does not know where in Europe John is, in our present context, it does not, because it is simply not relevant where in Europe John is. Unlike the disjunctive utterance, (9.37) is not significantly longer, and not more syntactically complex than more specific alternative utterances.

In summary, when a speaker utters a disjunction $p \vee q$, the following facts conspire to trigger an uncancelable, non-optional implicature:

1. There is a *ceteris paribus* preference for shorter, less complex forms, hence, everything else being equal, uttering p and uttering q is preferable to uttering $p \vee q$.

2. $p \vee q$ is asymmetrically entailed by p and by q , hence an utterance of p conveys the information that $p \vee q$ is true.
3. Because of 1. and 2., if the speaker wanted to convey $p \vee q$, and nothing prevented him from asserting p , he would have done so.
4. The speaker just uttered $p \vee q$ instead of p .

It follows that there is something that prevented the speaker from uttering p . In a cooperative dialogue, often the only plausible reason is that the speaker does not know that p is true, which is why in many contexts, the addressee can infer ignorance.

9.3.3 The NaR implicature of disjunction in dynamic pragmatics

I now show how the NaR implicature of disjunction arises in the framework of dynamic pragmatics. We need at least the following alternative utterances:³¹

$$(9.38) \quad \perp, \text{utter}(p), \text{utter}(q), \text{utter}(p \vee q)$$

It is of course crucial that p and q are considered as utterance alternatives. With Eckardt (2007), I think it is very natural to assume that both disjuncts are made salient by the utterance of the disjunction.³²

In terms of preferences, we already have encountered the crucial preference for shorter linguistic forms *MINIMIZE*, as well as the outcome preferences *QUALITY* and *INFORM*. Here I shall assume throughout that *INFORM* $p \vee q$ is present, i.e., that if $p \vee q$, the speaker prefers that the addressee know this. If we assumed also a preference *INFORM* p , then we would be in a variant of the scalar implicature scenario considered in Section 9.2.2.³³ Instead, we assume such a preference is absent. As

³¹I use the same notational simplifications as used in Section 9.2.

³²Interestingly, this poses some problems for classical theories which, like Gazdar (1979)'s, derive alternatives by a simple replacement of logical operators in the uttered sentence. Sauerland (2004) offers a technical fix for this problem, but it is tempting to simply give up the assumption that alternatives are derived by simple replacement.

³³The only difference would be that the weaker expression now incurs more violations on *MINIMIZE*, but that would not change decision outcomes.

discussed in the previous section, we still should predict that the implicature arises. And we do.

Deriving ignorance

In order to derive the ignorance implication (as opposed to more the general implication that the speaker ‘has a reason’), assume that *Ad* believes that the speaker has no other (relevant) preferences, i.e.,

$$(9.39) \quad \text{For all } v \in B_{Ad} : \text{EP}_{Sp,v} = \text{QUALITY} > \text{INFORM } p \vee q > \text{MINIMIZE}$$

In worlds in which *Sp* does not believe $p \vee q$ to be true, the optimal action is obviously \perp , as it is the only one that does not violate QUALITY. There are four logical possibilities for worlds in which *Sp* believes $p \vee q$ to be true, which differ in terms of which of p and q the speaker believes:

	$\Box_{Sp}p$	$\Box_{Sp}q$	
$v_{\neg\Box p \wedge \neg\Box q}$	FALSE	FALSE	Fig. 9.9
$v_{\Box p \wedge \neg\Box q}$	TRUE	FALSE	Fig. 9.10
$v_{\neg\Box p \wedge \Box q}$	FALSE	TRUE	Fig. 9.11
$v_{\Box p \wedge \Box q}$	TRUE	TRUE	Fig. 9.12

The outcome of **Opt** in each of these types of worlds is illustrated in Figs. 9.9 to 9.12, and summarized in (9.40).

$$(9.40) \quad \begin{array}{l} \text{a. } \text{Opt}(v_{\neg\Box p \wedge \neg\Box q}) = \{\text{utter}(p \vee q)\} \\ \text{b. } \text{Opt}(v_{\Box p \wedge \neg\Box q}) = \{\text{utter}(p)\} \\ \text{c. } \text{Opt}(v_{\neg\Box p \wedge \Box q}) = \{\text{utter}(q)\} \\ \text{d. } \text{Opt}(v_{\Box p \wedge \Box q}) = \{\text{utter}(p), \text{utter}(q)\} \end{array}$$

As a result $B_{Ad}[\text{utter}(p \vee q)]$ will only contain worlds of type $v_{\neg\Box p \wedge \neg\Box q}$ —that is, after observing the utterance of $p \vee q$, the addressee will believe that the speaker believes neither p nor q to be true. That is the ignorance implicature.

$w_{\neg p \wedge \neg q}$	QUALITY	INFORM $p \vee q$	MINIMIZE
\perp		*	
utter(p)	*		*
utter(q)	*		*
utter($p \vee q$)			***

Figure 9.9: Decision in worlds $v_{\neg p \wedge \neg q}$ where the speaker knows neither p nor q

$w_{p \wedge \neg q}$	QUALITY	INFORM $p \vee q$	MINIMIZE
\perp		*	
utter(p)			*
utter(q)	*		*
utter($p \vee q$)			***

Figure 9.10: Decision in worlds $v_{p \wedge \neg q}$ where the speaker knows p but not q

$w_{\neg p \wedge q}$	QUALITY	INFORM $p \vee q$	MINIMIZE
\perp		*	
utter(p)	*		*
utter(q)			*
utter($p \vee q$)			***

Figure 9.11: Decision in worlds $v_{\neg p \wedge q}$ where the speaker knows q but not p

$w_{p \wedge q}$	QUALITY	INFORM $p \vee q$	MINIMIZE
\perp		*	
utter(p)			*
utter(q)			*
utter($p \vee q$)			***

Figure 9.12: Decision in worlds $v_{p \wedge q}$ where the speaker knows both p and q

Alternative implicature: Unwillingness to inform

Ignorance is not the only reason why a speaker could opt for the disjunction instead of uttering either disjunct. He might also be unwilling to divulge which of p and q is true (while still communicating that one of them is true). We can account for this, too.

Let us consider an alternative addressee belief state B'_{Ad} in which the addressee is certain that the speaker believes either p or q , but not both, i.e.,

$$(9.41) \quad B_{Ad} \models \neg \Box_{Sp}(p \wedge q) \wedge (\Box_{Sp}(p) \vee \Box_{Sp}(q))$$

However, Ad does not yet have a belief as to which of p and q the speaker believes, and he takes it to be possible that Sp has the additional preference $\neg \text{INFORM } p/q$, which is satisfied if the addressee believes neither p nor q :

$$(9.42) \quad \neg \text{INFORM } p/q \\ \lambda v [v \models \neg \Box_{Ad} p \wedge \neg \Box_{Ad} q]$$

If the addressee is uncertain about (9.42), then there are four relevantly different worlds in B'_{Ad} (recall that we assume (9.41)):

	$\Box_{Sp} p$	$\Box_{Sp} q$	$\text{ep}(\neg \text{INFORM } p/q)$	
$v_{\Box p \wedge \text{ep}(\neg \text{INF})}$	TRUE	FALSE	TRUE	Fig. 9.13
$v_{\Box p}$	TRUE	FALSE	FALSE	Fig. 9.14
$v_{\Box q \wedge \text{ep}(\neg \text{INF})}$	FALSE	TRUE	TRUE	
$v_{\Box q}$	FALSE	TRUE	FALSE	

Figures 9.13 and 9.14 shows the outcome of Opt in worlds in which the speaker believes that p (the other two cases are symmetric). (9.43) summarizes the decision in the four kinds of worlds:

$$(9.43) \quad \begin{aligned} \text{a. } & \text{Opt}(v_{\Box p \wedge \text{ep}(\neg \text{INF})}) = \{\text{utter}(p \vee q)\} \\ \text{b. } & \text{Opt}(v_{\Box p}) = \{\text{utter}(p)\} \\ \text{c. } & \text{Opt}(v_{\Box q \wedge \text{ep}(\neg \text{INF})}) = \{\text{utter}(p \vee q)\} \\ \text{d. } & \text{Opt}(v_{\Box q}) = \{\text{utter}(q)\} \end{aligned}$$

$v_{\Box p \wedge ep(\neg \text{INF})}$	QUALITY	$\neg \text{INFORM } p/q$	$\text{INFORM } p \vee q$	MINIMIZE
\perp			*	
$\text{utter}(p)$		*		*
$\text{utter}(q)$	*			*
$\text{utter}(p \vee q)$				***

Figure 9.13: Decision in worlds where the speaker believes p and has a preference against revealing that

$v_{\Box p}$	QUALITY	$\text{INFORM } p \vee q$	MINIMIZE
\perp		*	
$\text{utter}(p)$			*
$\text{utter}(q)$	*		*
$\text{utter}(p \vee q)$			***

Figure 9.14: Decision in worlds where the speaker believes p and does not have a preference against revealing that

But then, updating B'_{Ad} with $\text{utter}(p \vee q)$ will only leave worlds in which the speaker has the preference $\neg \text{INFORM } p/q$. Thus we predict that if the addressee believes that the speaker knows which of p and q is true, but is uncertain whether the speaker is willing to share this information, observing an utterance of $p \vee q$ will let him conclude that the speaker prefers to withhold this information.

The general implicature

Of course, it may also be that the addressee's information state is a combination of the two we have just looked at. In that case, the addressee will remain uncertain as to whether the speaker does not know, or does not want to reveal, which of p and q is true. Again, this accords with intuition.

In general, as long as we assume that the preference MINIMIZE is present, an update with $\text{utter}(p \vee q)$ will only allow worlds v to survive such that

- (i) in v , the speaker has a preference C ;
- (ii) in v , C is ranked higher than MINIMIZE;

(iii) in v , Sp believes that $\text{utter}(p)$ violates C but $\text{utter}(p \vee q)$ does not.

That is, updating with $\text{utter}(S, p \vee q)$ removes all those worlds in which the speaker does not have a reason for not uttering p (and symmetrically for q). In the two cases we looked at, this defeating preference was either *QUALITY* (leading to the ignorance implicature) or \neg *INFORM* p/q (leading to an implicature that the speaker wants to withhold information). Of course, in the general case, these need not be the only two possible reasons. Here, as in the case of scalar implicatures in general, *any* kind of preference could lead the speaker to avoid the stronger alternative. Just as Grice's original theory has been amended by assuming, for example, a set of politeness maxims (Leech 1983, Brown and Levinson 1987), so in the present system we can assume speaker preferences that are motivated by politeness considerations. In this case, the reason an addressee may infer that the speaker used a disjunction for reasons of politeness, even though the speaker believes one of the disjuncts to be true and wishes the addressee believed the same. This is plausible for utterances of sentences like (9.44).

(9.44) Either you made a mistake or I did.

9.3.4 Generalizing NaR

We can now generalize the conditions under which NaR implicatures arise. Before doing so, I want to briefly note that the prediction of mandatory implicatures does not hinge on the particular choice of decision procedure made here. Indeed, *any* optimization-based analysis of implicatures potentially predicts such implicatures.

Here is why: As an analysis of implicatures, such a theory should allow us to characterize, for a given expression S and implicature i , the set $C_{S \rightarrow i}$ of contexts in which the implicature arises. At the same time, being an optimization-based theory, it allows us to characterize the set $C_{\text{Opt}(S)}$ of contexts in which S is optimal. Further, any such theory involves the assumption, on some level, that speakers utter S only in contexts that are in $C_{\text{Opt}(S)}$.³⁴ But then, the theory will predict i to be

³⁴At the very least, it seems, any such theory must attribute the belief that speakers only make

a mandatory implicature of S just in case (9.45) holds.

$$(9.45) \quad C_{\text{Opt}(S)} \subseteq C_{S \rightsquigarrow i}$$

That is, unless this situation is somehow excluded, any optimization-based theory has the potential to predict mandatory implicatures. Further, any optimization-based theory that assumes a general *ceteris paribus* preference for shorter or simpler forms will predict the NaR implicature of disjunction (provided that disjunctions count as longer or more complex in the relevant sense).

This is true, in particular, of the optimality and game-theoretic approaches referenced in the introduction to this chapter. These generally predict mandatory implicatures, though if their authors point out this fact, they seem to view it as a problem (perhaps due to the pervasive belief that all implicatures should be optional and cancelable).

As a concrete example, consider the game-theoretic account of Franke (2009, 2011). Franke (2011) presents the account in two stages. In the first stage, his game-models encode the assumption that the speaker is perfectly informed about the state of the world. As Franke discusses (p. 50–51), in these ‘base-level’ games, his analysis predicts that $p \vee q$ is not used.³⁵ It is only in the second stage, where he introduces ‘epistemically-lifted’ game models that can represent speaker-uncertainty, that his model predicts $p \vee q$ to be used, and to implicate uncertainty.³⁶ Of course, I think that this prediction is perfectly adequate. It captures the fact that the ‘ignorance’ implicature is mandatory.

In order to predict a mandatory implicature on any theory, including the one ‘optimal’ utterances to addressees, even if it allows for speakers to make non-‘optimal’ utterances. It is conceivable to have an optimization-based theory where this belief itself is an assumption that is only in place in some contexts—but then, this hypothesized theory of implicature really is only a theory of implicatures *for contexts in which this assumption is in place*. And for such contexts, the theory potentially predicts mandatory implicatures.

³⁵Technically, this prediction arises only if $p \wedge q$ is in the set of alternatives—without this alternative, his base-model in fact predicts, problematically, that $p \vee q$ is interpreted as $p \wedge q$.

³⁶Franke limits attention to situation in which (a) the addressee needs to know the state of the world exactly and (b) the preferences of the speaker and the hearer are perfectly aligned and (c) the preferences are mutually known. Presumably, if this assumption is lifted, Franke’s also could predict alternative NaR implicatures, e.g., that the speaker prefers to withhold information.

used in the present chapter, a number of additional prerequisites need to be in place. We can generalize from the case of disjunction as follows:

(9.46) An expression e will give rise to a NaR implicature if:

- a. **There is alternative expression e' which is informationally stronger than e .**

e' need not logically entail e , as it does in the case of disjunction. It can also be informationally stronger because of entrenched world-knowledge, for example.

- b. **e' is salient whenever e is uttered.**

In the case of disjunction, e' is actually a constituent of e , but this need not be the case, of course: All that is necessary is that observing an utterance of e makes e' salient.

- c. **There is a linguistic preference for uttering e' rather than e , all else being equal.**

In the case of disjunction, this preference plausible has to do with length or complexity, but this is not a necessity: Any preference between linguistic expressions will do.

For the implicature to be truly mandatory, of course, the preference in (9.46c) must be assumed to be universally present. It is here where an account of implicatures that relies on maxims of conversation and a cooperative principle (instead of a more general notion of a preference) may make different predictions. NaR implicatures are quite similar to classical implicatures that arise on the basis of the MAXIM OF MANNER, which after all also involves prescriptions about *how* things should be expressed. But the Gricean maxims, as classically understood, depend on the assumption of cooperativity. And indeed, that is why Grice thought that all implicatures must be cancelable. The quote I gave initially elided his reasoning:

“Since, to assume the presence of a conversational implicature, we have to assume that at least the Cooperative Principle is being observed, and since it is possible to opt out of the observation of this principle,

it follows that a [...] conversational implicature can be canceled in a particular case.”

(Grice 1975)

In a preference-based analysis of implicatures, we can construe the preference in (9.46c) instead as a *selfish* preference, which is in place even in cases in which the speaker cannot be taken to be cooperative. In this case, we predict truly mandatory implicatures. Without such an assumption of a selfish preference, we still predict *near*-mandatory implicatures, which arise in all contexts in which it is a given that the speaker is cooperative.

9.3.5 More NaR implicatures?

As pointed out in Section 9.3.1, the existence of NaR implicatures challenges widely-held pragmatic orthodoxy. Optionality and cancelability have long been used as a test for implicature-hood: In order to determine whether a given observed implication of a sentence is an implicature or not, it is checked whether the sentence can be uttered and, at the same time, the implication can be coherently denied. If this is not possible, it is concluded that the implication cannot be a conversational implicature, but must instead be semantic in kind. Indeed, Sadock (1978) called this ‘the best of the tests’ for implicature-hood. Once we realize that conversational implicatures can be mandatory, that means that using cancelability as a test is problematic: If it is used at all, it has to be used with great care.³⁷

On the plus side, this means that phenomena that may otherwise have seemed outside of the reach of pragmatic theory can possibly be accounted for in pragmatic terms after all, with the usual benefit of simplifying and streamlining semantic analyses. In this section, I want to briefly mention three possible cases that have

³⁷Hirschberg (1985)’s move to require cancelability as a property of implicatures by definition, of course, cures the symptom but not the illness: We *can* then use cancelability as a test for implicature-hood, but if a given implication is not cancelable, we still cannot conclude that it must be semantic in nature—because it could still be a non-implicature pragmatic implication.

been discussed in the literature in which NaR reasoning, or something close to it, might be at play.

Huitink and Spenader (2004)'s cancelation-resistant implicatures

The first case involves implicatures that are not mandatory in the sense that NaR implicatures are mandatory, but I want to discuss them for two reasons. On the one hand, these have been described as involving failures of cancelability, and I want to highlight the difference between the two kinds of non-cancelability. On the other hand, I think that NaR-style reasoning may be a better way to account for the robustness of these implicatures than the analysis that has been proposed.

Huitink and Spenader (2004) discuss cases were classical examples of implicatures are very difficult to cancel (Weiner (2006) discusses a similar case):

- (9.47) [In a letter of reference for a philosopher]
 Mr. X's command of English is excellent and his attendance at tutorials has been regular. He is a brilliant philosopher.
- (9.48) Miss X produced a series of sounds that corresponded closely with the score of "Home Sweet Home". She has a beautiful voice.

The first sentence in each of these examples is a classical example from Grice (1975): The first sentence in (9.47) is taken to implicate that Mr. X is no good in philosophy, while the first sentence in (9.48) is taken to implicate that Miss X sang badly. The second sentence in each case is the negation of the putative implicature.

As Huitink and Spenader note, it is actually quite difficult to take these sentences as cancelations: The much more natural interpretation is to re-interpret them as ironic statements and maintain the implicature.

Now, of course, even if it were *impossible* to interpret these examples as cancelations, this would not show much: At best, it shows that there are implicatures that cannot be canceled in every context in which they arise—obviously **Mr X's command of English is excellent** does not implicate **Mr X is a bad philosopher** in *every* context in which it is used. The NaR implicature of disjunction is mandatory

in a much stronger sense: There is *no* context where it is absent or can be canceled.

Huitink and Spenader propose that the implicatures in these cases are ‘semi-conventionalized’, by which they mean that there is a conventional link between a certain type of context and the implicature. While something like that *might* be plausible for (9.47), I find it hard to see how the same could be true for (9.48). In any case, NaR reasoning—of a more contextual sort than in the case of disjunction—can potentially help us to understand why the implicature is difficult to cancel in these contexts: In both cases (and the other cases Huitink and Spenader discuss), it is quite difficult to see why a speaker would say or write the first sentence if he believes the second.

Romero and Han (2004): High-negation polar questions

Romero and Han (2004) develop an account of high-negation polar questions (HNPs), such as (9.49).

(9.49) Doesn’t John drink?

↪ The speaker believes or at least expects that John drinks.

HNPs are an intricate topic which I do not wish to discuss here. I mention the issue because Romero and Han claim that the implication in (9.49) is an implicature, despite being ‘strong and uncancelable’—and they aim to account for it in a way that is very close to the NaR reasoning we have seen so far.

Han and Romero propose that HNPs contain a semantic operator that turns the question into a ‘meta-conversational move’ (essentially, a move having to do with common-ground management). Then they propose the following pragmatic principle (p. 629):

(9.50) Principle of Economy:

Do not use a meta-conversational move unless necessary (to resolve epistemic conflict or to ensure QUALITY).

It should be obvious how similar reasoning on the basis of such a principle is to NaR reasoning. Essentially, the principle says that a speaker must have a reason for making a ‘meta-conversational move’. The similarity is not merely superficial: Romero and Han take (9.50) to be an inviolable pragmatic principle (for otherwise, it could not explain why the epistemic implicature in (9.49) is uncancelable), but in my terms, we can reconstruct it by assuming a *ceteris paribus* preference against making ‘meta-conversational moves’. So, Romero and Han (2004)’s analysis of HNPQs fits very naturally into the general NaR picture.

Infelicity via implicature

The final case I want to discuss is perhaps the most promising one: NaR implicatures can be used to explain why certain sentences are *infelicitous*: If a sentence *S* that has a NaR implicature *i* is uttered in a context in which *i* is known to be false, it will be perceived infelicitous, because the addressee cannot make sense of the speaker’s utterance choice.

This is impossible with implicatures that are optional: If *S* only optionally implicates *i*, and *S* is used in a context in which *i* is known to be false, then *i* should simply be absent, instead of rendering the utterance infelicitous.

For this reason, a set of phenomena have recently been argued to involve grammaticalized scalar inferences (e.g., á la Chierchia, Fox and Spector (2012)),³⁸ on the grounds that the expressions triggering these inferences are infelicitous if the inference is known to be false (Magri 2009, Magri 2011, Ivlieva 2012). The argument has the general form in (9.51).

- (9.51) a. Sentences *s* of a given type are infelicitous, often to the point of appearing ungrammatical.
- b. Comparison with similar (felicitous) sentences suggests that *s* triggers a scalar inference *i*.

³⁸These authors maintain the use of the term ‘implicature’ for the grammaticalized inferences they hypothesize. I refer to them as ‘grammaticalized scalar inferences’, such inferences do not fit the usual conception of a scalar implicature.

- c. Based on logical entailment and/or entrenched world-knowledge, speakers know that *i* is incompatible with the truth conditions of *s*.
- d. The infelicity is hence due to the fact that the utterance of *s* triggers an inference that is known to be false if *s* is true.
- e. Because Gricean implicatures are always optional, *i* cannot be a Gricean implicature. Instead, *i* must be an inference arising as part of the semantic meaning of *s*.

We can accept steps (a-d) but deny (e) if *i* is a NaR implicature. The observed infelicity can then be explained in terms of pragmatic theory alone.

To illustrate: Magri (2009) examines individual-level predicates (i-predicates), which can be characterized as predicates that denote permanent properties. They contrast with stage-level predicates (s-predicates), which denote temporary properties. Since i-predicates behave differently from s-predicates in a number of ways, it had been proposed (e.g., by Kratzer (1995), Chierchia (1995)) that the two kinds of predicates differ in terms of their grammatical properties. Magri proposes that both kinds of predicates are grammatically the same, and aims to explain the observed differences in terms of obligatory scalar inferences. One such difference is that i-predicates are generally infelicitous (or ‘odd’) with various kinds of temporal modification, for example, temporal frame adverbials such as **this month**:

- | | | |
|--------|---|-------------|
| (9.52) | a. John is of noble birth. | i-predicate |
| | b. John is in in financial trouble. | s-predicate |
| (9.53) | a. #John is of noble birth this month. | i-predicate |
| | b. John is in financial trouble this month. | s-predicate |

Magri’s explanation for the infelicity of (9.53a) goes roughly as follows: The sentences in (9.53) give rise to the obligatory scalar inferences in (9.54). However, due to world knowledge, it will generally be common ground that if John is of noble birth this month, he is always of noble birth. But then, the scalar inference (9.54a) triggered by (9.53a) is contextually incompatible with the asserted content

of the sentence. The utterance gives rise to contradictory implications, and hence is infelicitous.

- (9.54) a. John is not of noble birth at times before or after this month.
 b. John is not in financial trouble at times before or after this month.

For this explanation to work, the scalar inference cannot be optional: If it were, (9.53a) should be felicitous, because the offending inference should simply be absent. Magri concludes that the scalar inference hence cannot be a Gricean implicature, and instead proposes to derive it within Chierchia et al. (2012)'s grammatical theory of implicature.

Magri's account is attractive, as it obviates the need to represent the distinction between *i*- and *s*-predicates in the grammar. The existence of Gricean NaR implicatures has the potential of simplifying the explanation even further, predicting the observed contrasts based on a uniform semantic representation and pragmatic reasoning. The idea, essentially, is this: Given entrenched the world-knowledge that 'being of noble birth' is a permanent property, the (longer, hence dispreferred) sentence **John is of noble birth this month** is *informationally equivalent* to the (shorter, hence preferred) sentence **John is of noble birth**. But then, there cannot be a reason to utter the dispreferred form instead of the preferred one. Consequently, the utterance of the dispreferred form is perceived as odd.

Of course, for this explanation to be convincing, we have to establish that the prerequisites for NaR implicatures are met. We have to establish that **John is of noble birth** is preferred (due to length, complexity, or other considerations) to **John is of noble birth this month**, and that any utterance of the latter sentence makes the former sentence salient as an alternative. But both of these seem to be sensible assumptions, at least at first glance.

It is worth stressing what the success of such an account would mean: It would mean that we can explain why a given expression is *never* felicitous (at least as we maintain the required world-knowledge assumptions), on entirely Gricean grounds. Such explanations have long been thought to be beyond the reach of Gricean theory, which was construed, after all, as a theory concerned with weak,

defeasible, cancelable inferences. NaR reasoning shows that this conception is incorrect, and it flows from the general conviction that formed the starting point of this dissertation: Gricean reasoning is operational on *every* occasion of language use.

Chapter 10

Outlook

This dissertation has developed a formal framework for pragmatics that is faithful to the Gricean conception of pragmatic inference as interlocutors' reasoning about utterance choice. It has emphasized the usefulness of such a Gricean perspective for formal semantics and pragmatics beyond its classical field of application, i.e., the study of non-entailed inferences. The framework incorporates an articulated theory of sentential force, and allows fine-grained modeling of the interaction between semantic content, force, interactional reasoning and contextual assumptions.

In the present chapter, I want to reprise some of the open questions that were raised, discuss some of the shortcomings of the current framework and simplifications that were made, and make some brief remarks about how these issues could be addressed.

10.1 The role of intentions in pragmatic theory

The framework of dynamic pragmatics does not represent *intentions*. As I pointed out in Chapter 1, this may seem surprising for a pragmatic theory that calls itself 'Gricean', considering the central role Grice saw for intentions in determining speaker meaning.

Intentions vs. effective preferences

One reason for this choice was that ‘intention’, at least as it is understood in philosophy, is a rather intricate and complicated concept. I have opted to use a novel technical term, ‘effective preferences’, to talk about non-epistemic attitudes that influence action choice. Looking at things this way, the concept of effective preference may be viewed as ‘intention lite’.

However, it is not clear that everything that I want to consider an effective preference—a preference that shapes the speaker’s action choice—can sensibly be called an intention. One example are ‘linguistic preferences’ such as the preference for shorter, simpler forms that has played a considerable role in Chapter 9. While it makes sense to assume that the utterance choices of speakers are affected by such a preference, it is not quite clear if it makes sense to call it an intention. I therefore think it is preferable to view of intentions as just *one* of the various preferential attitudes that can figure as effective preferences of an agent.¹

The relevance of intentions

At the same time, a number of the analyses presented in the course of this dissertation did not even assume an effective preference in cases where intentions are commonly assumed to play a crucial role. The analysis of how addressees come to believe in the truth of the content (Chapters 2 and 5) did not assume that (the addressee believes that) the speaker has an effective preference to communicate the content of his utterance. Similarly, according to the analysis presented in Chapter 7, it does not matter whether a speaker who utters **I hereby promise to be there** has an effective preference for committing himself (or for being there): His utterance will be a promise (and commit him) regardless. Finally, in Chapter 9, I showed how an implicature can arise without assuming that (the addressee believes that) the speaker effectively prefers to communicate it.

¹Though perhaps, intentions are given a greater weight than other such attitudes. This would go some way to explain the ‘special role’ intentions (as opposed to mere desires) play in our decision making according to Bratman (1987), as well as some of the properties Bratman requires intentions to have—viz., that they be consistent.

In each of these cases, of course, a speaker *could* have such an effective preference, and, in certain context, it may be relevant whether he does and whether the addressee recognizes that he does. But the basic phenomena seem to be the same regardless of whether such an effective preference is present.

Perhaps the investigation of other phenomena will reveal that intentions play a more central role in them. In the meantime, I take it to be an open question whether intentions, and their recognition, figure in a central way in our understanding of everyday language use.

10.2 Ambiguity and underspecification

At the very outset of this dissertation, Chapter 2, I made a radically idealizing assumption. I assumed that agents observe each other's utterances perfectly, and went on to ignore all kinds of ambiguity and underspecification, essentially assuming (in modeling speakers of a simple propositional language) that agents utter disambiguated logical forms, with all contextual parameters filled in.

Ultimately, we want to lift this assumption. This is not only because we know that, as a matter of fact, natural languages are rife with underspecification and ambiguity (at multiple levels of linguistic description). It also is likely that in resolving such ambiguity and underspecification, agents employ the same kind of Gricean reasoning as I have described in the course of this dissertation. Two options for treating ambiguity and underspecification in the current framework naturally suggest themselves:

Ambiguous utterance events. The first option is to change the nature of the utterance events we have assumed, but hold on to the assumption that utterance events are perfectly observed. Instead of having events $\text{utter}(i, i', \varphi)$, where φ is mapped to a unique interpretation (e.g., a set of possible worlds), we would have $\text{utter}(i, i', S)$, where S is mapped to a *set* of interpretations (e.g., a set of sets of possible worlds).

This would require an adjustment in our conventions of use. Recall that the

utterances of many sentences create *speaker commitments* to beliefs or preferences. We would have to specify what it means to be committed to a belief in a possibly-ambiguous sentence, rather than a proposition.

Imperfect perception of utterance events. The second option is to maintain that utterance events are individuated by the disambiguated logical form, but give up the assumption that events are perfectly perceived. For a sentence S that is ambiguous between two readings φ_1 and φ_2 , we would have two utterance types:

- (10.1) a. $\text{utter}(i, i', S, \varphi_1)$
 b. $\text{utter}(i, i', S, \varphi_2)$

Then we have to replace the PAL constraint to reflect that, if (10.1a) happens, a hearer learns that *either* (10.1a) *or* (10.1b) happened.

A promising option for finding a suitable new constraint on belief change is to turn to generalizations of the models for PUBLIC ANNOUNCEMENT LOGIC that inspired our PAL constraint. In particular, the ‘event models’ of Baltag, Moss and Solecki (1998)² allow for just this kind of selective indistinguishability of events. van Benthem (2011, Chapter 4) gives an accessible introduction to the basic framework, which has been employed to interpret various systems of multi-agent DYNAMIC EPISTEMIC LOGIC (DEL).

Transition Preference Pragmatics. Either of these options, if implemented appropriately, would lead to a model that is quite close to a combination of the dynamic pragmatics developed here and the TRANSITION PREFERENCE PRAGMATICS (TPP) of Beaver (2002).

Beaver employs a dynamic system in which updates can be *indeterminate*, and hence an addressee may need to employ pragmatic reasoning in order to identify the correct transition between contexts. Beaver himself employs this system

²Baltag et al. (1998) called these models ‘action models’, but in the more recent literature, the name ‘event models’ has become standard.

mainly for modeling ambiguity and underspecification, but it has also been employed by Davis (2011) to model the effect of sentence-final particles and intonation in Japanese—both devices that affect the conventional effect of the uttered sentence. As Davis otherwise uses a conception of the basic effects of the main clause types that is very similar to the one in this dissertation, integrating something like TPP into the current framework offers the possibility of unifying his theoretical perspective with the one taken in this dissertation.

10.3 A question of commitment

Any treatment of ambiguity and underspecification in the current framework raises a theoretical question: How does the commitment induced by an ambiguous utterance get determined? If we allow for ambiguous utterance events, this question manifests itself in the issue of how the conventions of use should be adjusted to take into account that utterance events do not determine a single interpretation. If instead we maintain that utterance events determine their contents (but allow for hearer-uncertainty about which utterance happened), the question manifests itself in the issue of how the event that *actually* happens in a world is determined.

In either case, the question comes down to this: When a speaker utters (10.2), what determines whether he becomes committed to the belief that Cleo is German, or the belief that Hanne is German, or . . . ?

(10.2) She is German.

In Gricean terms, this question amounts to how ‘what is said’ gets determined—in this example, what determines who **she** refers to? The classical answer involves appeal to speaker intentions: **She** refers to whichever individual the speaker intended to refer to. Perhaps intentions will make their great entrance here; perhaps *this* is where intentions are crucial in the current view of things. They determine what the speaker refers to with an underspecified referential expression, and how ambiguities get resolved. Thereby, intentions determine what the speaker gets

committed to by his utterance.

But once again, I am sceptical that it is a speaker's private intentions that are the deciding factor. Conceptually, private intentions just do not mesh with the very public nature of commitments. A speaker who utters (10.2) cannot freely choose what he gets committed to by having a private intention to refer to some arbitrary (female) individual, without publicizing this intention somehow, and he cannot avoid becoming committed at all by secretly failing to have a referential intention.

I leave this as an open question. The current framework brings out sharply that this *is* a question, and one that is distinct, in principle, from the question how audiences figure out what the speaker wanted to communicate. The question **What did the speaker really, truly refer to?** may seem like an idle one in cases in which the speaker had a certain interpretation in mind, and the addressee correctly identified this interpretation—but in the current set-up, this question has a very practical bite to it: The answer determines what the speaker becomes committed to, and hence, for example, what kind of future utterances the speaker can make without retraction.

10.4 Belief revision, salience, and awareness

Throughout this thesis, I have often made rather strong assumptions about agents' beliefs. For example, I have occasionally assumed that an addressee believes, with certainty, that the speaker has a certain effective preference. This is often implausible in practice, which makes it tempting to add a representation of *graded belief* (e.g., subjective probability) to the model. This is a possibility, but there are also a number of related issues whose solution may solve the problem of unreasonably strong assumptions about belief at the same time:

There is a tension between two properties of the framework as introduced in this dissertation: (i) The framework does not properly model belief revision, as such revision is essentially unconstrained; (ii) the framework does not incorporate a treatment of salience, attention, or limited awareness of possibilities.

It is defensible to ignore the issue of belief revision, if one conceptualizes the

modeled beliefs in an appropriate way. We can say that the \Box_i -operator models very strongly-held beliefs, which agents form cautiously and rarely have occasion to give up. Then we can maintain that revision of *these* beliefs happens so rarely that we can often ignore the issue in practice.

Conversely, we can make do without an explicit model of salience or limited awareness if we assume that beliefs are very context-dependent. To model a situation in which an agent only takes salient possibilities into account, we simply can assume that the speaker believes that the salient possibilities are the only ones. But if we do this, belief revision is a common occurrence, and we need to model it faithfully.

One way to resolve this tension is to extend the current system in such a way that it can account for agents being unaware of possibilities which their beliefs, in principle, do not exclude. A particularly promising approach is the model of Franke and de Jager (2011).³ They employ a standard Kripke model as a specification of ‘background beliefs’, which gets ‘filtered’ through a state of limited awareness. The result of the filtering operation is again a standard Kripke model, but one in which certain possibilities are ignored. Modalities and decision procedures are then defined in the usual way on this ‘foreground’ model.

This kind of approach is promising for two reasons: Firstly, it means that we can deal with issues such as salience without changing the basic set-up of our models—we simply add awareness states and specify their dynamics. Secondly, explicitly modeling (lack of) awareness to possibilities is a very intuitive way to model the effect of utterances that do not actually provide novel information, but instead serve as reminders that make a certain possibility salient, as in Franke and de Jager’s (10.3), and similar examples that they discuss.

- (10.3) [A is looking for his keys.]
 B: Could they be in the car?

³As Franke and de Jager (2011, Section 5), discuss, their model is inspired by models of ‘unawareness’ in computer science and rational choice theory (e.g., Fagin and Halpern (1988), Feinberg (2004)), but differs in the details, in part because Franke and de Jager are mainly interested in a kind of unawareness that is *easily overturned*—a property that is also desirable for our purposes here.

So amending the model with a treatment of unawareness (or similar models of ‘default’ beliefs, such as Veltman’s (1996) treatment of defaults in update semantics) will allow us to apply the current framework to more phenomena. At the same time, doing so could remedy, in one fell swoop, the two shortcomings of the system mentioned above: The lack of a proper model of belief revision becomes defensible, as we can conceptualize the ‘background’ beliefs as ‘cautious’ beliefs that rarely need to be revised. At the same time, we can model instances in which only a limited range of possibilities is taken into account (e.g., the assumptions about utterance alternatives discussed in Section 9.1.2), without making unreasonable assumptions about what the agents believe to be possible.

10.5 Conclusion

I have only scratched the surface of the range of phenomena that can be illuminated by systematically taking a Gricean perspective in the way the system of dynamic pragmatics does. In addition to putting forth particular analyses, and independently from the particulars of the formal framework, I hope this dissertation serves as an advertisement for this kind of perspective on language use. Thinking about language use—*every* occasion of language use—as an instance of purposive human behavior sheds new light on old questions, makes us see new generalizations, and raises new questions.

Appendix A

The basic system

This appendix contains the basic definitions for the system set up in Chapter 2.

Definition 1 (*Prop syntax*). *Given a set of proposition letters P , the language $Prop$ is the smallest set such that:*

- (i) $P \subseteq Prop$.
- (ii) If $\varphi \in Prop$ then so is $\neg\varphi$.
- (iii) If $\varphi_1, \varphi_2 \in Prop$, then so are $(\varphi_1 \wedge \varphi_2)$, $(\varphi_1 \vee \varphi_2)$, $(\varphi_1 \rightarrow \varphi_2)$, $(\varphi_1 \leftarrow \varphi_2)$ and $(\varphi_1 \leftrightarrow \varphi_2)$.

Definition 2 (*Prop interpretation*). *Given an a set of worlds W and an interpretation function $I : P \mapsto \wp(W)$, $\llbracket \cdot \rrbracket^{Prop}$ is defined as:*

- For $p \in P : \llbracket p \rrbracket^{Prop} = I(p)$.
- $\llbracket (\phi \wedge \psi) \rrbracket^{Prop} = \llbracket \phi \rrbracket^{Prop} \cap \llbracket \psi \rrbracket^{Prop}$.
- $\llbracket \neg\phi \rrbracket^{Prop} = W \setminus \llbracket \phi \rrbracket^{Prop}$.

For other formulas, $\llbracket \cdot \rrbracket^{Prop}$ is defined using the usual equivalences.

Definition 3 (*Pragmatic Language: Syntax*). *For a given object language $Prop$, we define \mathcal{P}_{Prop} (the PRAGMATIC LANGUAGE FOR $Prop$) as follows: Let*

- (a) I_c, T_c, E_c be sets of constant symbols (the ‘individual constants’, the ‘temporal constants’ and the ‘event constants’, respectively).
- (b) I_v, T_v, E_v be sets of variable symbols (the ‘individual variables’, the ‘temporal variables’ and the ‘event constants’, respectively).
- (c) \mathbb{P} be a set of predicate symbols and ar a function $\mathbb{P} \mapsto \mathbb{N}$.

such that $I_c, T_c, I_v, T_v, E_c, E_v, \mathbb{P}$ are pairwise disjoint. Then \mathcal{P}_{Prop} is the smallest set such that:

1. If $\varphi \in Prop$, then $\varphi \in \mathcal{P}_{Prop}$.
2. If $\phi, \psi \in \mathcal{P}_{Prop}$, so are $\neg\phi, (\phi \wedge \psi), (\phi \vee \psi), (\phi \rightarrow \psi), (\phi \leftarrow \psi)$ and $(\phi \leftrightarrow \psi)$.
3. If $\phi \in \mathcal{P}_{Prop}$ and $i \in I_t \cup I_v, t \in T_c \cup T_v$, then $\Box_{i,t}\phi \in \mathcal{P}_{Prop}$.
4. If $\phi \in \mathcal{P}_{Prop}$ and $t \in T_c \cup T_v$, then $C_t\phi \in \mathcal{P}_{Prop}$.
5. If $\phi \in \mathcal{P}_{Prop}$ and $x \in I_v \cup T_v \cup E_v$, then $\exists_x : \phi \in \mathcal{P}_{Prop}$.
6. If $P \in \mathbb{P}$ and $ar(P) = n$ and $a_1, \dots, a_n \in I_c \cup I_v \cup L$ and $t \in T_c \cup T_v$, then $P_t(a_1, \dots, a_n) \in \mathcal{P}_{Prop}$.
7. If $\varphi \in L, i, i' \in I, e \in E_c \cup E_v$, then
 - a. $utter_e(i, i', \ulcorner \varphi \urcorner) \in \mathcal{P}_{Prop}$.
 - b. $utter_e(i, \ulcorner \varphi \urcorner) \in \mathcal{P}_{Prop}$.
 - c. $utter_e(i) \in \mathcal{P}_{Prop}$.

Definition 4 ($T \times W$ frames). A $T \times W$ frame is a quadruple $\langle W, T, <, \approx \rangle$, where W and T are non-empty sets, $<$ is a transitive relation on T which is also irreflexive and linear, and \approx is a 3-place relation on $T \times W \times W$, such that (1) for all t , \approx_t is an equivalence relation and (2) for all $w_1, w_2 \in W$ and $t, t' \in T$, if $w_1 \approx_t w_2$ and $t' < t$ then $w_1 \approx_{t'} w_2$. (Thomason 1984, Definition 6, p. 216)

Definition 5 ($\mathbb{N} \times W$ frames). A $\mathbb{N} \times W$ frame is a $T \times W$ frame $\langle W, T, <, \approx \rangle$ where $T = \mathbb{N}$ and $<$ has its usual interpretation.

Definition 6 (\mathcal{P}_{Prop} frames). A \mathcal{P}_{Prop} frame is a $\langle W, \approx, Ind, Ag, R \rangle$, where $\langle W, \approx \rangle$ is a $\mathbb{N} \times W$ frame, Ind a domain of individuals, $Ag \subseteq Ind$ a set of agents, and R a function $Ag \times \mathbb{N} \mapsto \wp(W \times W)$ that takes agent, time pairs into relations between worlds, such that

- $R_{i,t}$ is transitive, serial, and Euclidean for all i, t ;
- *historicity*: If $w_1 \approx_t w_2$, then $w_1 R_{i,t} v$ iff $w_2 R_{i,t} v$.
- *no fore-belief*: If $v_1 \approx_t v_2$, then $w R_{i,t} v_1$ iff $w R_{i,t} v_2$.

Definition 7 (Transitive Closure). For a given \mathcal{P}_{Prop} -frame $\langle W, \approx, Ind, Ag, R \rangle$, let $\mathbb{R} : \mathbb{N} \mapsto W \times W$ be the function such that for every $t \in \mathbb{N}$, $\mathbb{R}(t)$ is the smallest relation such that:

- (i) $\bigcup_{i \in Ind} R_{i,t} \subseteq \mathbb{R}(t)$.
- (ii) $\mathbb{R}(t)$ is transitive.

Definition 8 (Event types). Let \mathbb{E} be a set of event classes, Ind a set of individuals, and L a language. Then the set of event types based on \mathbb{E}, Ind, L is defined as:

$$Ev(\mathbb{E}, Ind, L) := \{P(i_1, \dots, i_n) \mid P \in \mathbb{E} \ \& \ i_1, \dots, i_n \in Ind \cup L\}$$

Definition 9 (\mathcal{P}_{Prop} models). Given a language L , a model \mathcal{M} for \mathcal{P}_{Prop} is a tuple $\langle \mathcal{F}, I, \mathbb{E}, Hap \rangle$ where

- (i) $\mathcal{F} = \langle W, \approx, Ind, Ag, R \rangle$ is a \mathcal{P}_{Prop} frame
- (ii) I is an interpretation function such that
 - a. $I(t) \in \mathbb{N}$ if $t \in T_c$.
 - b. $I(i) \in Ind$ if $i \in I_c$.
 - c. $I(e) \in T \times T$ if $e \in E_c$.

- d. $I(P) \in W \mapsto T \mapsto (Ind \cup Prop)^n$ if $P \in \mathbb{P}$ and $ar(P) = n$.
- e. $I(p) \in \wp(W)$ if p is a proposition letter of $Prop$.

(iii) \mathbb{E} is a set of event classes such that $\mathbf{utter} \in \mathbb{E}$.

(iv) $\mathbf{Hap} : W \times (T \times T) \mapsto Ev(\mathbb{E}, Ind, L)$ such that

- a. for any w , $\mathbf{Hap}_w(t, t')$ is defined only if $t' = t + 1$.
- b. if $w_1 \approx_t w_2$ then for all $t_1, t_2 \leq t$: $\mathbf{Hap}_{w_1}(t_1, t_2) = \mathbf{Hap}_{w_2}(t_1, t_2)$ (historicity).
- c. if $w_1 \approx_t w_2$ and $\mathbf{Hap}_{w_1}(t, t+1) = \mathbf{Hap}_{w_2}(t, t+1)$ then $w_1 \approx_{t+1} w_2$ (determinism).

Additionally, we require that I respects \approx , that is that for $P \in \mathbb{P}$: $I(P)(w, t) = I(P)(w', t)$ if $w \approx_t w'$.

Definition 10 (\mathcal{P}_{Prop} satisfaction). For a given language $Prop$, given a \mathcal{P}_{Prop} -model $\langle \langle W, \approx, Ind, Ag, R \rangle, I, \mathbb{E}, \mathbf{Hap} \rangle$, and an assignment g, \vDash is the following relation between W and \mathcal{P}_{Prop} (where I_g is the function such that $I_g(x) = g(x)$ if x is a variable, $I_g(x) = I(x)$ otherwise):

1. $w \vDash^s \varphi$ if $\varphi \in Prop$ and $w \in \llbracket \varphi \rrbracket^{Prop}$.
2. $w \vDash^s P_t(a_1, \dots, a_n)$ iff $\langle I_g(a_1), \dots, I_g(a_n) \rangle \in I(w)(I_g(t))(P)$.
3. $w \vDash^s \Box_{i,t} \phi$ if for all $v : wR_{I_g(i), I_g(t)} v : v \vDash^s \phi$.
4. $w \vDash^s C_t \phi$ if for all $v : wR(I_g(t))v : v \vDash^s \phi$.
5. $w \vDash^s \exists_x : \phi$ if there is $d \in Ind \cup T \cup (T \times T) : w \vDash^{s[x/d]} \phi$.
6. a. $w \vDash^s \mathbf{utter}_e(a, b, \ulcorner \varphi \urcorner)$ iff $\mathbf{Hap}_w(I_g(e)) = \mathbf{utter}(I_g(a), I_g(b), \varphi)$.
 b. $w \vDash^s \mathbf{utter}_e(a, \ulcorner \varphi \urcorner)$ iff there is $i \in Ind : \mathbf{Hap}_w(I(e)) = \mathbf{utter}(I_g(a), i, \varphi)$.
 c. $w \vDash^s \mathbf{utter}_e(a)$ iff there are $i \in Ind, \varphi \in L : \mathbf{Hap}_w(I_g(e)) = \mathbf{utter}(I_g(a), i, \varphi)$.
7. Propositional connectives are interpreted in the usual way.

Bibliography

- Ajdukiewicz, K.: 1938, *Logiczne Podstawy Nauczania [The logical foundation of teaching]*, Nasza Księgarnia, Warsaw/Vilnius.
- Ariel, M.: 2004, Most, *Language* **80**(4), 658–706.
- Atlas, J. and Levinson, S.: 1981, It-clefts, informativeness and logical form, in P. Cole (ed.), *Radical Pragmatics*, Academic Press, New York, NY, pp. 1–61.
- Aumann, R.: 1995, Backward induction and common knowledge of rationality, *Games and Economic Behavior* **8**(1), 6–19.
- Aumann, R.: 1996, Reply to Binmore, *Games and Economic Behavior* **17**(1), 138–146.
- Austin, J. L.: 1962, *How to do things with words*, Harvard University Press, Cambridge, MA.
- Bach, K.: 2006, The top ten misconceptions about implicature, in B. Birner and G. Ward (eds), *Drawing the Boundaries of Meaning: Neo-Gricean Studies in Pragmatics and Semantics in Honor of Laurence R. Horn*, John Benjamins, Amsterdam/New York, NY, pp. 21–30.
- Bach, K.: 2012, Saying, meaning and implicating, in K. Allan and K. M. Jaszczolt (eds), *Cambridge Handbook of Pragmatics*, Cambridge University Press, Cambridge, UK, pp. 47–68.
- Bach, K. and Harnish, R. M.: 1979, *Linguistic Communication and Speech Acts*, MIT Press, Cambridge, MA.

- Bach, K. and Harnish, R. M.: 1992, How performatives really work: A reply to Searle, *Linguistics and Philosophy* **15**(1), 93–110.
- Baltag, A., Moss, L. S. and Solecki, S.: 1998, The logic of public announcements, common knowledge, and private suspicions, in I. Gilboa (ed.), *Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK) 7*, Morgan Kaufmann Publishers, San Francisco, CA, pp. 43–56.
- Beaver, D.: 2001, *Presupposition and Assertion in Dynamic Semantics*, CSLI Publications, Stanford, CA.
- Beaver, D. I.: 2002, Pragmatics, and that's an order, in D. Barker-Plummer, D. I. Beaver, J. van Benthem and P. S. di Luzi (eds), *Words, Proofs and Diagrams*, CSLI Publications, Stanford, CA, pp. 191–216.
- Belnap, N. and Perloff, M.: 1990, Seeing to it that: A canonical form for agentives, in H. Kyburg, R. Loui and G. Carlson (eds), *Knowledge representation and defeasible reasoning*, Vol. 5 of *Studies in Cognitive Systems*, Kluwer Academic Publishers, Dordrecht, pp. 167–190.
- Benz, A. and van Rooij, R.: 2007, Optimal answers, and what they implicate, *Topoi* **26**(1), 63–78.
- Bierwisch, M.: 1980, Semantic structure and illocutionary force, in J. R. Searle, F. Kiefer and M. Bierwisch (eds), *Speech Act Theory and Pragmatics*, Reidel, Dordrecht, pp. 1–35.
- Blutner, R.: 1998, Lexical Pragmatics, *Journal of Semantics* **15**(2), 115–162.
- Blutner, R.: 2000, Some aspects of optimality in natural language interpretation, *Journal of Semantics* **17**(3), 189–216.
- Blutner, R.: 2002, Lexical semantics and pragmatics, *Linguistische Berichte, Sonderheft* **10**, 27–58.
- Boghossian, P. A.: 1989, The rule-following considerations, *Mind* **98**(392), 507–549.

- Brandom, R.: 1983, Asserting, *Noûs* 17(4), 637–650.
- Bratman, M. E.: 1987, *Intention, Plans, and Practical Reason*, Harvard University Press, Cambridge, MA.
- Bresnan, J.: 2000, Optimal syntax, in J. Dekkers, F. van der Leeuw and J. van de Weijer (eds), *Optimality Theory: Phonology, Syntax and Acquisition*, Oxford University Press, Oxford, UK, pp. 334–385.
- Brown, P. and Levinson, S.: 1987, *Politeness: Some universals in language usage*, Cambridge University Press, Cambridge, UK.
- Castroviejo Miró, E.: 2008, Deconstructing exclamations, *Catalan Journal of Linguistics* 7, 41–90.
- Chernilovskaya, A., Condoravdi, C. and Lauer, S.: 2012, On the discourse effects of *wh*-exclamatives, in N. Arnett and R. Bennett (eds), *Proceedings of the 30th West Coast Conference on Formal Linguistics (WCCFL)*, Cascadilla Press, Somerville, MA, pp. 109–119.
- Chernilovskaya, A., Condoravdi, C. and Lauer, S.: in prep., The context change effect of exclamatives. Manuscript, Stanford and Utrecht University.
- Chernilovskaya, A. and Nouwen, R.: 2012, On *wh*-exclamatives and noteworthiness, in M. Aloni, F. Roelofsen, G. W. Sassoon, K. Schulz, N. Kimmelman and M. Westera (eds), *Proceedings of the Amsterdam Colloquium 2011*, Springer, Berlin.
- Chierchia, G.: 1995, Individual-level predicates as inherent generics, in G. N. Carlson and F. J. Pelletier (eds), *The generic book*, University of Chicago Press, Chicago, IL, pp. 125–175.
- Chierchia, G., Fox, D. and Spector, B.: 2012, Scalar implicature as a grammatical phenomenon, in C. Maienborn, K. Heusinger and P. Portner (eds), *Semantics. An International Handbook of Natural Language Meaning*, Vol. 3, De Gruyter

- Mouton, pp. 2297–2332.
(Previous versions were circulated under the title *The Grammatical View of Scalar Implicatures and the Relationship between Semantics and Pragmatics*).
- Chierchia, G. and McConnell-Ginet, S.: 1990, *Meaning and Grammar: An introduction to semantics*, MIT Press, Cambridge, MA.
- Ciardelli, I. and Roelofsen, F.: 2011, Inquisitive logic, *Journal of Philosophical Logic* 40(1), 55–94.
- Condoravdi, C.: 2002, Temporal interpretation of modals: Modals for the present and for the past, in D. I. Beaver, S. Kaufmann, B. Clark and L. Casillas (eds), *The Construction of Meaning*, CSLI Publications, Stanford, CA, pp. 59–88.
- Condoravdi, C. and Lauer, S.: 2009, Performing a wish: Desiderative assertions and performativity. Talk presented at California Universities Semantics and Pragmatics (CUSP) 2, UC Santa Cruz, November 2009.
- Condoravdi, C. and Lauer, S.: 2011, Performative verbs and performative acts, in I. Reich, E. Horch and D. Pauly (eds), *Sinn and Bedeutung 15: Proceedings of the 2010 Annual Conference of the Gesellschaft für Semantik*, Universaar – Saarland University Press, Saarbrücken, pp. 149–164.
- Condoravdi, C. and Lauer, S.: 2012, Imperatives: Meaning and illocutionary force, in C. Piñón (ed.), *Empirical Issues in Syntax and Semantics* 9, pp. 37–58.
- Condoravdi, C. and Lauer, S.: to appear, Anankastic conditionals are just conditionals, *Proceedings of Semantics and Linguistic Theory (SALT) 23*, Cornell University, CLC Publications, Ithaca, NY.
- Davis, C.: 2009, Decisions, dynamics, and the Japanese particle *yo*, *Journal of Semantics* 26(4), 329–366.
- Davis, C.: 2011, *Constraining Interpretation: Sentence Final Particles in Japanese*, PhD thesis, University of Massachusetts at Amherst.

- Degen, J. and Franke, M.: 2012, Optimal reasoning about referential expressions, in S. Brown-Schmidt, J. Ginzburg and S. Larsson (eds), *Proceedings of SeineDial*, Université Paris–Diderot (Paris 7), Paris Sorbonne-Cité, Paris, France, pp. 2–11.
- Doerge, F.: 2009, A scholarly confusion of tongues, or, is promising an illocutionary act?, *Lodz Papers in Pragmatics* 5(1), 53–68.
- Douven, I.: 2006, Assertion, knowledge, and rational credibility, *Philosophical Review* 115(4), 449–485.
- Eckardt, R.: 2007, Licensing ‘or’, in S. Uli and P. Stateva (eds), *Presupposition and Implicature in Compositional Semantics*, Palgrave Macmillan, New York, NY, pp. 34–70.
- Fagin, R. and Halpern, J.: 1988, Belief, awareness and limited reasoning, *Artificial Intelligence* 34(1), 39–76.
- Faller, M.: 2002, *Semantics and Pragmatics of Evidentials in Cuzco Quechua*, PhD thesis, Stanford University.
- Farkas, D. F. and Bruce, K. B.: 2010, On Reacting to Assertions and Polar Questions, *Journal of Semantics* 27(1), 81–118.
- Farkas, D. F. and Roelofsen, F.: forthcoming, Polar initiatives and polarity particles in an inquisitive discourse model, *Language* .
- Feinberg, Y.: 2004, Subjective reasoning—Games with unawareness, *Research Paper 1875*, Stanford Graduate School of Business, Stanford, CA.
- Fox, D. and Katzir, R.: 2011, On the characterization of alternatives, *Natural Language Semantics* 19(1), 87–107.
- Frank, M. C. and Goodman, N. D.: 2012, Predicting pragmatic reasoning in language games, *Science* 336(6084), 998.

- Franke, M.: 2008, Meaning and inference in case of conflict, in K. Balogh (ed.), *Proceedings of the 13th ESSLLI Student Session*, pp. 65–74.
- Franke, M.: 2009, *Signal to Act: Game Theory and Pragmatics*, PhD thesis, Universiteit van Amsterdam.
- Franke, M.: 2011, Quantity implicatures, exhaustive interpretation, and rational conversation, *Semantics and Pragmatics* 4(1), 1–82.
- Franke, M. and de Jager, T.: 2011, Now that you mention it: Awareness dynamics in discourse and decisions, in A. Benz, C. Ebert, G. Jäger and R. Rooij (eds), *Language, Games, and Evolution*, Vol. 6207 of *Lecture Notes in Computer Science*, Springer, Berlin/Heidelberg, pp. 60–91.
- Franke, M., de Jager, T. and van Rooij, R.: 2012, Relevance in cooperation and conflict, *Journal of Logic and Computation* 22(1), 23–54.
- García-Carpintero, M.: 2013, Explicit performatives revisited, *Journal of Pragmatics* 49, 1–17.
- Gazdar, G.: 1979, *Pragmatics: Implicature, Presupposition and Logical Form*, Academic Press, New York.
- Giddens, A.: 1984, *The Constitution of Society*, University of California Press, Berkeley, CA.
- Ginet, C.: 1979, Performativity, *Linguistics and Philosophy* 3(2), 245–265.
- Ginzburg, J.: 1996, Interrogatives: Questions, facts and dialogue, in S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*, Blackwell Textbooks in Linguistics, Blackwell, pp. 385–422.
- Grice, H. P.: 1957, Meaning, *The Philosophical Review* 66(3), 377–388.
- Grice, H. P.: 1969, Utterer's meaning and intention, *The Philosophical Review* 78(2), 147–177.

- Grice, H. P.: 1975, Logic and conversation, in P. Cole and J. L. Morgan (eds), *Speech Acts*, Vol. 3 of *Syntax and Semantics*, Academic Press, pp. 41—58.
- Grice, H. P.: 1978, Further notes on logic and conversation, in P. Cole (ed.), *Syntax and Semantics 9: Pragmatics*, Academic Press, New York.
- Grice, H. P.: 1982, Meaning revisited, in N. V. Smith (ed.), *Mutual Knowledge*, Academic Press. page numbers refer to the reprint in (Grice 1989).
- Grice, H. P.: 1989, *Studies in the Way of Words*, Harvard University Press.
- Grimshaw, J.: 1979, Complement selection and the lexicon, *Linguistic Inquiry* 10(2), 279–326.
- Groenendijk, J. and Roelofsen, F.: 2009, Inquisitive semantics and pragmatics, in J. M. Larrazabal and L. Zubeldia (eds), *Meaning, Content, and Argument: Proceedings of the ILCLI International Workshop on Semantics, Pragmatics, and Rhetoric*, University of the Basque Country, Publication Service, pp. 41–72.
- Groenendijk, J. and Stokhof, M.: 1984, *Studies on the Semantics of Questions and the Pragmatics of Answers*, PhD thesis, Universiteit van Amsterdam.
- Groenendijk, J. and Stokhof, M.: 1991, Dynamic predicate logic, *Linguistics and Philosophy* 14(1), 39–100.
- Groenendijk, J., Stokhof, M. and Veltman, F.: 1995, Coreference and modality, in S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*, Blackwell Publishers, Oxford, UK, pp. 179–213.
- Gunlogson, C.: 2003, *True to Form: Rising and Falling Declaratives in English*, Routledge, New York.
- Gunlogson, C.: 2008, A question of commitment, *Belgian Journal of Linguistics* 22(1), 101–136.

- Gutiérrez-Rexach, J.: 1996, The semantics of exclamatives, in E. Garrett and F. Lee (eds), *Semantics at sunset. UCLA working papers in linguistics*, University of California, Los Angeles, Los Angeles, CA, pp. 146–162.
- Hamblin, C. L.: 1958, Questions, *Australasian Journal of Philosophy* **36**(3), 159–168.
- Hamblin, C. L.: 1971, Mathematical models of dialogue, *Theoria* **37**(2), 130–155.
- Han, C.-H.: 2002, Interpreting interrogatives as rhetorical questions, *Lingua* **112**(3), 201–229.
- Hansson, S. O.: 1996, What is ceteris paribus preference?, *Journal of Philosophical Logic* **25**(3), 307–332.
- Hare, R. M.: 1968, Wanting: Some pitfalls, in R. Binkley (ed.), *Agent, Action and Reason, Proceedings of the Western Ontario Colloquium*.
- Hare, R. M.: 1970, Meaning and speech acts, *Philosophical Review* **79**(1), 3–24.
- Harnish, R. M.: 2005, Commitments and speech acts, *Philosophica* **75**(1), 11–41.
- Hausser, R.: 1978, Surface compositionality and the semantics of mood, in J. Groenendijk and M. Stokhof (eds), *Amsterdam Papers in Formal Grammar*, Vol. II, Universiteit van Amsterdam.
- Heal, J.: 1974, Explicit performative utterances and statements, *The Philosophical Quarterly* **24**(95), 106–121.
- Hedenius, I.: 1963, Performatives, *Theoria* **29**(2), 115–136.
- Heim, I.: 1983, File change semantics and the familiarity theory of definiteness, in R. Bäuerle, R. Schwarze and A. von Stechow (eds), *Meaning, Use and the Interpretation of Language*, Walter de Gruyter, Berlin, pp. 164–190.
- Hendriks, P. and de Hoop, H.: 2001, Optimality theoretic semantics, *Linguistics and Philosophy* **24**(1), 1–32.

- Hindriks, F.: 2009, Constitutive rules, language, and ontology, *Erkenntnis* 71(2), 253–275.
- Hintikka, J.: 1961, Modality and quantification, *Theoria* 27(3), 119–128.
- Hirschberg, J.: 1985, *A theory of scalar implicature*, PhD thesis, University of Pennsylvania.
- Hiz, H.: 1978, Introduction, in H. Hiz (ed.), *Questions*, Reidel, Dordrecht.
- Horn, L. R.: 1972, *On the Semantic Properties of Logical Operators in English*, PhD thesis, University of California, Los Angeles.
- Horn, L. R.: 1984, Towards a new taxonomy of pragmatic inference: Q-based and R-based implicature, *Meaning, Form, and Use in Context: Linguistic Applications*, Georgetown University Press, Washington, DC, pp. 11–42.
- Horn, L. R.: 1989, *A Natural History of Negation*, Chicago University Press, Chicago, IL.
- Horn, L. R.: 2006, The border wars: A neo-Gricean perspective, in K. von Stechow and K. Turner (eds), *Where semantics meets pragmatics*, Vol. 16 of *Current Research in the Semantics/Pragmatics Interface*, Elsevier, Amsterdam/London, pp. 21–48.
- Huitink, J. and Spenader, J.: 2004, Cancellation resistant PCis, in R. van der Sandt and B. Geurts (eds), *Proceedings of the ESSLLI 2004 Workshop on Implicature and Conversational Meaning*.
- Ivlieva, N.: 2012, Obligatory implicatures and grammaticality, in M. Aloni, V. Kimmelman, F. Roelofsen, G. Sassoon, K. Schulz and M. Westera (eds), *Logic, Language and Meaning. Proceedings of the 18th Amsterdam Colloquium, Revised Selected Papers*, Vol. 7218 of *Lecture Notes in Computer Science*, Springer, Berlin/Heidelberg, pp. 381–390.
- Jäger, G.: 2007, Game dynamics connects semantics and pragmatics, in A.-V. Pietarinen (ed.), *Game Theory and Linguistic Meaning*, Elsevier, Amsterdam/New York, pp. 89–102.

- Jäger, G.: 2012, Game theory in semantics and pragmatics, in C. Maienborn, P. Portner and K. von Stechow (eds), *Semantics. An International Handbook of Natural Language Meaning*, Vol. 3, De Gruyter Mouton, Berlin, pp. 2487–2516.
- Jäger, G. and Ebert, C.: 2009, Pragmatic rationalizability, in A. Riester and T. Solstad (eds), *Proceedings of Sinn und Bedeutung 13*, University of Stuttgart, OPUS.
- Joshi, A. K.: 1982, Mutual belief in question answering systems, in N. S. Smith (ed.), *Mutual Knowledge*, Academic Press, London, pp. 181–197.
- Kadmon, N.: 2001, *Formal Pragmatics: Semantics, Pragmatics, Presupposition, and Focus*, Blackwell, Malden, MA.
- Kamp, H.: 1978, Semantics versus pragmatics, in F. Guenther and S. J. Schmidt (eds), *Formal Semantics and Pragmatics for Natural Languages*, Reidel, Dordrecht, pp. 255–287.
- Kamp, J. A. W.: 1981, A theory of truth and semantic representation, in J. Groenendijk, T. Janssen and M. Stokhof (eds), *Formal Methods in the Study of Language*, Vol. 135 of *Mathematical Centre Tracts*, Mathematisch Centrum, Amsterdam, pp. 277–322.
- Kanger, S.: 1957, *Provability in Logic*, Almqvist and Wiksell, Stockholm.
- Kaplan, D.: 1999, What is meaning? Explorations in the theory of *meaning as use*. Manuscript, Brief Version – Draft # 1. University of California, Los Angeles.
- Karttunen, L.: 1977, Syntax and semantics of questions, *Linguistics and Philosophy* 1(1), 3–44.
- Katzipir, R.: 2007, Structurally-defined alternatives, *Linguistics and Philosophy* 30(6), 669–690.
- Kaufmann, M.: 2012, *Interpreting Imperatives*, Springer, Dordrecht/New York.
- Kaufmann, S.: 2005, Conditional truth and future reference, *Journal of Semantics* 22(3), 231–280.

- Kaufmann, S. and Schwager, M.: 2009, A uniform analysis of conditional imperatives, in E. Cormany, S. Ito and D. Lutz (eds), *Proceedings of Semantics and Linguistic Theory (SALT) 19*, Cornell University, CLC Publications, Ithaca, NY, pp. 239–256.
- Kratzer, A.: 1981, The notional category of modality, in H. J. Eikmeyer and H. Rieser (eds), *Words, Worlds, and Contexts. New Approaches in Word Semantics*, de Gruyter, Berlin, pp. 38–74.
- Kratzer, A.: 1995, Stage-level and individual-level predicates, in G. N. Carlson and F. J. Pelletier (eds), *The generic book*, The University of Chicago Press, Chicago, pp. 125–175.
- Krifka, M.: 2001a, For a structured meaning account of questions and answers, in C. Féry and W. Sternefeld (eds), *Audiatur Vox Sapientiae: A Festschrift for Arnim von Stechow*, Akademie-Verlag, Berlin, pp. 287–319.
- Krifka, M.: 2001b, Quantifying into question acts, *Natural Language Semantics* 9(1), 1–40.
- Krifka, M.: to appear, Embedding speech acts, in T. Roeper and P. Speas (eds), *Recursion in Language and Cognition*.
- Kripke, S.: 1963, Semantical analysis of modal logic, I. normal modal propositional calculi, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9, 67–96.
- Lackey, J.: 2007, Norms of assertion, *Noûs* 41(4), 594–626.
- Landman, F.: 1986, *Towards a Theory of Information: The status of partial objects in semantics*, Foris Publications, Dordrecht/Riverton, MJ.
- Laserson, P.: 1999, Pragmatic halos, *Language* 75(3), 522–551.
- Lauer, S.: 2012, On the pragmatics of pragmatic slack, in A. Aguilar, R. Nouwen and A. Chernilovskaya (eds), *Proceedings of Sinn and Bedeutung (SuB) 16*, MIT Working papers in Linguistics, Cambridge, MA, pp. 389–401.

- Lauer, S. and Condoravdi, C.: 2012, The basic dynamic effect of interrogative utterances. Talk presented at the 13th Texas Linguistics Society (TLS) conference, University of Texas, Austin, June 2012.
- Leech, G. N.: 1983, *Principles of Pragmatics*, Longman, London/New York.
- Lemmon, E. J.: 1962, On sentences verifiable by their use, *Analysis* **22**(4), 86–89.
- Levinson, D.: 2003, Probabilistic model-theoretic semantics for *want*, in R. Young and Y. Zhou (eds), *Proceedings of Semantics and Linguistic Theory (SALT) 13*, Cornell University, CLC Publications, Ithaca, NY, pp. 222–239.
- Levinson, S.: 1983, *Pragmatics*, Cambridge University Press, Cambridge, UK.
- Levinson, S.: 2000, *Presumptive Meanings*, MIT Press, Cambridge, MA.
- Lewis, D.: 1969, *Convention: A philosophical study*, Harvard University Press, Cambridge, MA.
- Lewis, D.: 1975, Languages and language, in K. Gunderson (ed.), *Language, Mind, and Knowledge*, University of Minnesota Press, Minneapolis, MN, pp. 3–35.
- Lewis, D.: 1979, Scorekeeping in a language game, *Journal of Philosophical Logic* **8**(1), 339–359.
- Lewis, D. and Lewis, S. R.: 1975, Review of Olson and Paul (1972), *Theoria* **41**(1), 39–60.
- MacFarlane, J.: 2005, Making sense of relative truth, *Proceedings of the Aristotelian Society* **105**(1), 305–323.
- MacFarlane, J.: 2011, What is assertion?, in J. Brown and H. Cappelen (eds), *Assertion*, Oxford University Press, Oxford, UK, pp. 79–96.
- Magri, G.: 2009, A theory of individual-level predicates based on blind mandatory scalar implicatures, *Natural Language Semantics* **17**(3), 245–297.

- Magri, G.: 2011, Another argument for embedded scalar implicatures based on oddness in downward entailing contexts, *Proceedings of Semantics and Linguistic Theory (SALT) 20*, Cornell University, CLC Publications, Ithaca, NY, pp. 564–581.
- Mastop, R.: 2005, *What can you do?*, PhD thesis, Universiteit van Amsterdam.
- Mayol, L. and Castroviejo, E.: 2013, How to cancel an implicature, *Journal of Pragmatics* **50**(1), 84–104.
- Olson, R. and Paul, A.: 1972, *Contemporary Philosophy in Scandinavia*, Johns Hopkins, Baltimore/London.
- Owens, D.: 2006, Testimony and assertion, *Philosophical Studies* **130**(1), 105–129.
- Parikh, P.: 2001, *The Use of Language*, CSLI Publications, Stanford, CA.
- Portner, P.: 2005, The semantics of imperatives within a theory of clause types, in R. B. Young (ed.), *Proceedings of Semantics and Linguistic Theory (SALT) 14*, Vol. 4, Cornell University, CLC Publications, Ithaca, NY, pp. 235–252.
- Portner, P.: 2007, Imperatives and modals, *Natural Language Semantics* **15**, 351–383.
- Portner, P.: 2012, Permission and choice, in G. Grewendorf and T. E. Zimmermann (eds), *Discourse and Grammar: From Sentence Types to Lexical Categories*, Studies in Generative Grammar, De Gruyter Mouton, Berlin, pp. 43–68.
- Potts, C.: 2005, *The Logic of Conventional Implicatures*, Oxford University Press, Oxford, UK.
- Potts, C.: 2007, The expressive dimension, *Theoretical Linguistics* **33**(2), 165–197.
- Potts, C. and Kawahara, S.: 2004, Japanese honorifics as emotive definite descriptions, in K. Watanabe and R. B. Young (eds), *Proceedings of Semantics and Linguistic Theory (SALT) 14*, Cornell University, CLC Publications, Ithaca, NY, pp. 235–254.

- Prince, A. and Smolensky, P.: 1993, Optimality theory: Constraint interaction in generative grammar, *Technical Report 2*, Rutgers University Center for Cognitive Science.
- Ransdell, J.: 1971, Constitutive rules and speech-act analysis, *The Journal of Philosophy* 68(13), 385–400.
- Åqvist, L.: 1965, *A new approach to the logical theory of interrogatives*, University of Uppsala.
- Åqvist, L.: 1983, On the “tell me truly” approach to the analysis of interrogatives, in F. Kiefer (ed.), *Questions and Answers*, Reidel, Dordrecht, pp. 9–14.
- Rett, J.: 2008, A degree account of exclamatives, in T. Friedman and S. Ito (eds), *Proceedings of SALT XVIII*, Cornell University, Ithaca, NY, pp. 601–618.
- Rett, J.: 2011, Exclamatives, degrees and speech acts, *Linguistics and Philosophy* 34(5), 411–442.
- Roberts, C.: 1996, Information structure in discourse: Towards an integrated account of formal pragmatics, *Papers in Semantics, OSU Working Papers in Linguistics*, Vol. 49, Department of Linguistics, The Ohio State University, Columbus, OH.
- Rohde, H.: 2006, Rhetorical questions as redundant interrogatives, *San Diego Linguistics Papers* 2(134–168).
- Romero, M. and Han, C.-H.: 2004, On negative “yes/no” questions, *Linguistics and Philosophy* 27(5), 609–658.
- Ross, J. R.: 1970, On declarative sentences, in R. Jacobs and P. Rosenbaum (eds), *Readings in English transformational grammar*, Ginn, Waltham, MA, pp. 222–272.
- Russell, B.: 2006, Against grammatical computation of scalar implicatures, *Journal of Semantics* 23(4), 361–382.

- Sadock, J. M.: 1978, On testing for conversational implicature, *Syntax and Semantics 9: Pragmatics*, Academic Press, New York, pp. 281–297.
- Sadock, J. M. and Zwicky, A. M.: 1985, Speech act distinctions in syntax, in T. Shopen (ed.), *Language Typology and Syntactic Description*, Vol. I, Cambridge University Press, Cambridge, UK, pp. 155–196.
- Sauerland, U.: 2004, Scalar implicatures in complex sentences, *Linguistics and Philosophy* **27**(3), 367–391.
- Schlenker, P.: 2010, Presuppositions and local contexts, *Mind* **119**(474), 377–391.
- Schmerling, S.: 1982, How imperatives are special and how they aren't, in R. Schneider, K. Tuite and R. Chametzky (eds), *Papers from the Parasession on Nondeclaratives: Chicago Linguistic Society*, Chicago, IL, pp. 202–218.
- Schwager, M.: 2006, *Interpreting Imperatives*, PhD thesis, Johann Wolfgang Goethe-Universität, Frankfurt am Main.
- Searle, J. R.: 1969, *Speech Acts: An essay in the philosophy of language*, Cambridge University Press, Cambridge, UK.
- Searle, J. R.: 1989, How performatives work, *Linguistics and Philosophy* **12**(5), 535–558.
- Siebel, M.: 2003, Illocutionary acts and attitude expression, *Linguistics and Philosophy* **26**(3), 351–366.
- Sperber, D. and Wilson, D.: 1986, *Relevance: Communication and Cognition*, Blackwell Publishers, Oxford, UK.
- Stalnaker, R.: 1974, Pragmatic presupposition, in M. Munitz and P. Unger (eds), *Semantics and Philosophy*, New York University Press, New York, pp. 197–213.
- Stalnaker, R.: 1978, Assertion, in P. Cole (ed.), *Syntax and Semantics 9: Pragmatics*, New York Academic Press, New York, pp. 315–332.

- Stalnaker, R.: 1994, On the evaluation of solution concepts, *Theory and Decision* **37**, 49–74.
- Stalnaker, R.: 1996, Knowledge, belief and counterfactual reasoning in games, *Economics and Philosophy* **12**(2), 133–163.
- Stalnaker, R.: 2002, Common ground, *Linguistics and Philosophy* **25**(5-6), 701–721.
- Stenius, E.: 1967, Mood and language-game, *Synthese* **17**, 254–274.
- Strawson, P. F.: 1964, Intention and convention in speech acts, *The Philosophical Review* **73**(4), 439–460.
- Szabolcsi, A.: 1982, Model theoretic semantics of performatives, in F. Kiefer (ed.), *Hungarian and General Linguistics*, John Benjamins, Amsterdam, pp. 515–536.
- Thomason, R. H.: 1984, Combinations of tense and modality, in D. Gabbay and F. Guenther (eds), *Handbook of Philosophical Logic: Extensions of Classical Logic*, Reidel, Dordrecht, pp. 135–165.
- van Benthem, J.: 2011, *Logical Dynamics of Information and Interaction*, Cambridge University Press, Cambridge, UK/New York.
- van Benthem, J. and Pacuit, E.: 2006, The tree of knowledge in action: Towards a common perspective, in Guido Governatori, I. Hodkinson and Y. Venema (eds), *Advances in Modal Logic*, Vol. 6, College Publications, London.
- van Ditmarsch, H., van der Hoek, W. and Kooi, B. (eds): 2008, *Dynamic Epistemic Logic*, Vol. 337 of *Studies in Epistemology, Logic, Methodology and Philosophy of Science*, Springer, Dordrecht.
- van Rooij, R.: 2004, Signaling games select Horn strategies, *Linguistics and Philosophy* **27**(4), 493–527.
- Veltman, F.: 1996, Defaults in update semantics, *Journal of Philosophical Logic* **25**(3), 221–261.

- Vogel, A., Potts, C. and Jurafsky, D.: 2013, Implicatures and nested beliefs in approximate Decentralized-POMDPs, *Proceedings of the 2013 Annual Conference of the Association for Computational Linguistics*, Association for Computational Linguistics, Stroudsburg, PA, pp. 74–80.
- von Savigny, E.: 1988, *The social foundations of meaning*, Springer, Berlin.
- Weiner, M.: 2006, Are all conversational implicatures cancellable?, *Analysis* **66**(2), 127–130.
- Williams, B.: 2002, *Truth and Truthfulness*, Princeton University Press, Princeton, NJ.
- Williamson, T.: 1996, Knowing and asserting, *Philosophical Review* **105**(4), 489–523.
- Witek, M.: 2009, Scepticism about reflexive intentions refuted, *Lodz Papers in Pragmatics* **5**(1), 69–83.
- Zanuttini, R. and Portner, P.: 2003, Exclamative clauses: At the syntax-semantics interface, *Language* **79**(1), 39–81.
- Zimmermann, M.: 2004, Zum *wohl*: Diskurspartikeln als Satztypmodifikatoren, *Linguistische Berichte* **199**, 253–286.
- Zimmermann, T. E.: 2000, Free choice disjunction and epistemic possibility, *Natural Language Semantics* **8**(4), 255–290.