

File ID	uvapub:52782
Filename	Schulz.pdf
Version	unknown

---

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type	PhD thesis
Title	Minimal models in semantics and pragmatics : free choice, exhaustivity, and conditionals
Author(s)	K. Schulz
Faculty	FNWI: Institute for Logic, Language and Computation (ILLC)
Year	2007

FULL BIBLIOGRAPHIC DETAILS:

<http://hdl.handle.net/11245/1.272471>

---

*Copyright*

*It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content licence (like Creative Commons).*

---

# Minimal Models in Semantics and Pragmatics

Free Choice, Exhaustivity, and Conditionals

ILLC Dissertation Series DS-2007-04



INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation

Universiteit van Amsterdam

Plantage Muidersgracht 24

1018 TV Amsterdam

phone: +31-20-525 6051

fax: +31-20-525 5206

e-mail: [illc@science.uva.nl](mailto:illc@science.uva.nl)

homepage: <http://www.illc.uva.nl/>

# Minimal Models

## in Semantics and Pragmatics

### Free Choice, Exhaustivity, and Conditionals

ACADEMISCH PROEFSCHRIFT

*ter verkrijging van de graad van doctor aan de  
Universiteit van Amsterdam  
op gezag van de Rector Magnificus  
prof. dr. D.C. van den Boom  
ten overstaan van een door het college voor  
promoties ingestelde commissie, in het openbaar  
te verdedigen in de Aula der Universiteit  
op vrijdag 2 november 2007, te 12.00 uur*

*door*

Katrin Schulz,  
geboren te Berlijn,  
Bondsrepubliek Duitsland

Promotiecommissie

Promotor: prof. dr. F.J.M.M Veltman

Co-promotor: dr. P.J.E. Dekker

Overige leden:

prof. dr. J.A.G. Groenendijk

dr. M. Aloni, postdoc

prof. dr. H.E. de Swart

prof. dr. N. Asher

prof. dr. C. Condoravdi

Faculteit der Geesteswetenschappen

The investigations were supported by the Netherlands Organization for Scientific Research (NWO), division Humanities (GW).

Copyright chapter 2 © 2005 by Springer

Copyright chapter 3 © 2006 by Springer

Copyright except the chapters 2 and 3 © 2007 by Katrin Schulz

Cover design by PRINTPARTNERS IPSKAMP and Katrin Schulz.

Cover photograph by Katrin Schulz.

Printed and bound by PRINTPARTNERS IPSKAMP, Enschede.

ISBN: 90-5776-164-5

---

# Contents

<b>Acknowledgments</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 The paradox of free choice permission</b>	<b>7</b>
2.1 Introduction . . . . .	7
2.2 Free choice inferences . . . . .	11
2.3 The approach . . . . .	13
2.3.1 Introduction . . . . .	13
2.3.2 The semantics . . . . .	14
2.3.3 Introducing the general ideas . . . . .	16
2.3.4 Working out the details . . . . .	17
2.3.4.1 The epistemic case . . . . .	18
2.3.4.2 The deontic case . . . . .	21
2.3.4.3 Competence . . . . .	22
2.3.4.4 Solving the paradox of free choice permission . . . . .	23
2.3.5 The cancellation of free choice inferences . . . . .	26
2.3.6 Conclusions . . . . .	27
2.4 Discussion . . . . .	28
2.4.1 An open problem . . . . .	28
2.4.2 Comparison . . . . .	29
2.4.2.1 The approaches of Kamp and Zimmermann . . . . .	29
2.4.2.2 Gazdar's approach to clausal implicatures . . . . .	30
2.5 Conclusions . . . . .	33
<b>3 Exhaustive interpretation</b>	<b>37</b>
3.1 Introduction . . . . .	37
3.2 The phenomenon . . . . .	39
3.2.1 Interaction with the semantic meaning of the answer . . . . .	39

3.2.2	The context-dependence of exhaustivity . . . . .	40
3.2.3	Other types of questions . . . . .	42
3.3	Groenendijk and Stokhof's proposal . . . . .	44
3.4	Exhaustivity as Predicate Circumscription . . . . .	47
3.4.1	Predicate Circumscription . . . . .	47
3.4.2	The basic setting . . . . .	49
3.5	Exhaustivity and dynamic semantics . . . . .	51
3.6	Exhaustivity and relevance . . . . .	56
3.6.1	The indirect approach . . . . .	58
3.6.2	The direct approach . . . . .	59
3.7	Exhaustive interpretation as conversational implicature . . . . .	61
3.8	Conclusion and outlook . . . . .	69
<b>4</b>	<b>Conditional sentences</b>	<b>73</b>
4.1	Introduction . . . . .	73
4.2	Central ideas . . . . .	75
4.3	Terminological preliminaries . . . . .	77
4.4	Caveat lector . . . . .	81
<b>5</b>	<b>The meaning of the conditional connective</b>	<b>83</b>
5.1	Introduction . . . . .	83
5.2	The similarity approach to conditionals . . . . .	85
5.3	Similarity as similarity of the past . . . . .	86
5.3.1	Backtracking counterfactuals . . . . .	87
5.3.2	The future similarity objection . . . . .	93
5.4	Premise semantics . . . . .	95
5.4.1	A short history of premise semantics . . . . .	95
5.4.2	Explaining Mr. Jones with premise semantics . . . . .	98
5.4.3	Problems of the approach . . . . .	100
5.5	Counterfactuals in causal networks . . . . .	102
5.5.1	The general ideas . . . . .	102
5.5.2	The formalization . . . . .	103
5.5.3	More examples . . . . .	109
5.5.4	Discussion . . . . .	112
5.6	Two readings for conditionals . . . . .	121
5.6.1	Motivation . . . . .	121
5.6.2	The epistemic reading . . . . .	129
5.6.2.1	Formalization . . . . .	131
5.6.2.2	Discussion of the epistemic reading . . . . .	135
5.6.3	The ontic reading . . . . .	139
5.6.3.1	Formalization . . . . .	140
5.6.3.2	Discussion of the ontic reading . . . . .	146
5.6.4	Discussion . . . . .	152

5.7	Summary . . . . .	156
<b>6</b>	<b>Tense in English conditionals</b>	<b>159</b>
6.1	Introduction . . . . .	159
6.2	The puzzle of the missing interpretation . . . . .	161
6.2.1	The observations . . . . .	161
6.2.2	Past-as-past approaches . . . . .	164
6.2.3	Past-as-modal approaches . . . . .	175
6.2.3.1	The past-as-unreal hypothesis . . . . .	176
6.2.3.2	The past-as-metaphor hypothesis . . . . .	179
6.2.3.3	The past-as-relict hypothesis . . . . .	180
6.2.3.4	The life-cycle hypothesis . . . . .	181
6.3	The puzzle of the shifted temporal perspective . . . . .	183
6.3.1	The observations . . . . .	184
6.3.2	Approaches to the observations . . . . .	191
6.3.3	Summary . . . . .	198
6.4	The proposal . . . . .	199
6.4.1	An introduction . . . . .	199
6.4.2	The language . . . . .	201
6.4.3	The model . . . . .	207
6.4.4	The interpretation of the vocabulary of $\mathcal{L}$ . . . . .	211
6.4.4.1	The epistemic update with atomic formulas. . . . .	213
6.4.4.2	The ontic update with atomic formulas. . . . .	214
6.4.4.3	Support and enforcement . . . . .	226
6.4.4.4	The meaning of the basic logical operators . . . . .	227
6.4.4.5	The meaning of the temporal operators. . . . .	228
6.4.4.6	The meaning of the perfect . . . . .	229
6.4.4.7	The meaning of the modals . . . . .	230
6.4.4.8	The meaning of the moods . . . . .	235
6.4.4.9	The meaning of $IF$ . . . . .	247
6.5	Discussion . . . . .	253
6.6	Summary . . . . .	269
<b>A</b>	<b>Appendix to chapter 5</b>	<b>273</b>
<b>B</b>	<b>Appendix to chapter 6</b>	<b>275</b>
	<b>Bibliography</b>	<b>277</b>
	<b>Index</b>	<b>289</b>
	<b>Samenvatting</b>	<b>293</b>





---

## Acknowledgments

First, I would like to thank my promotores Frank Veltman and Paul Dekker. Frank Veltman I thank particularly for sharing his ideas with me. They have inspired in many ways my thinking on conditionals. I am very grateful to Paul Dekker for the care with which he read my work on conditionals, even though the topic of this research does not stand central in his own work.

Furthermore, I would like to thank all my teachers, who have encouraged me – in different ways – to follow my interests and my curiosity, wherever they lead me. Let me mention some of the teachers I am particularly indebted to. Frank Beckman started my interests in formal semantics and pointed me to Stuttgart, from which it was only a small step to Amsterdam. Rainer Bäuerle supervised my stay in Stuttgart and suggested conversational implicatures as research topic. My fascination for implicatures started then and has never stopped since. It lead, via my Diplom thesis and my Master thesis, to the work reported on in the second and the third chapter of the present book. Another teacher I am grateful to is Hans Kamp, who set with his own example my standards for what good semantic work is. Finally, I also want to thank Michiel van Lambalgen for offering me a Ph.D. position in Amsterdam and for having faith in my abilities, even though I tried to convince him of the contrary.

I consider myself very lucky with having had the opportunity to do my Ph.D. at the ILLC in Amsterdam. This institute provides a highly inspiring but also very heartily and warm environment to work in. I am thankful to all my colleagues at the ILLC, particularly those on the second floor of the Philosophy department for making me love my work.

I am indebted to a number of people for comments on earlier versions of the second and the third chapter of the thesis. I should especially mention Maria Aloni, Luis Alonso-Ovalle, Jeroen Groenendijk, Benjamin Spector, Martin Stokhof and a number of anonymous reviewers.

Thomas Icard III deserves my warmest thanks for correcting the English of the thesis – of course, he is not responsible for any mistakes that are still present.

Robert van Rooij I have to thank in many ways. His work on formal pragmatics has been a great inspiration for my look on this area. He has been a wonderful supervisor when I wrote my Diplom thesis. Afterwards he became an excellent colleague and co-worker. The third chapter of this thesis is a result of this cooperation.

Finally, I would like to thank Robert, Simon and Hanna for making me happy.

Amsterdam  
August, 2007.

Katrin Schulz

## Chapter 1

---

# Introduction

It is a common truth that the simplest and most obvious questions are often particularly difficult to answer. For instance, when meeting new people I get confronted with the very reasonable but disturbing question *What are you doing for work?*. I normally answer that I do research in linguistics and hope that this will stop all further inquiries. Sometimes, this does not work and the questioner continues asking what is it that I am investigating. That is where the real trouble starts. The most correct answer would probably be that I am studying meaning, more precisely, the meaning of expressions of natural languages like English. But what is this meaning? There are so many facets to it, so many ways to look at meaning that it is hardly possible to give a satisfying and compact answer to this question. Already the very vague description just given raises a lot of questions. Is it the meaning of words I am considering or the meaning of sentences, for words seem to mean different things in different sentences? Maybe also the level of sentences is not abstract enough. Even for people that have never consciously thought about meaning before it is obvious that sentences mean different things under different circumstances. A sentence like *This is my husband* may be meant purely to tell the addressee which of the persons in a room is the husband of the speaker. But uttered in a bar to some fellow making you pretty uneasy, you may actually intend to communicate *Leave me alone*. Or, when you utter the sentence pointing to your dog, you certainly do not mean it to be true in a strict sense. You probably just want to express that you have (in some respects) the kind of relationship with your dog that married women normally have with their husband. Given that the meaning of a sentence depends on its actual use, do we, therefore, rather have to consider the meaning of a concrete occurrence of a sentence? But then we would not be able to account for those aspects of meaning common to all uses of a word or a sentence.

Such considerations have lead to a fundamental distinction in linguistics between the meaning of an expression by itself and the meaning that an expression can

obtain through interaction with the context in which it is used. With Grice (1989) one distinguishes two subtypes of meaning. There is, first, *semantic* meaning – as Grice puts it: *what is said*. This is the meaning carried by the words themselves. But there is also *pragmatic* meaning, meaning based on rules governing the use of a particular expression with its semantic meaning. Both semantic and pragmatic meaning together are taken to constitute the meaning of an expression. Semantic meaning is commonly described – at least until the early 80-ties – using truth conditions. Theories for pragmatic meaning are less uniform, but it is very popular to describe this part of meaning using theories of rational behavior such as decision theory and game theory.

However, the picture is still not as clear as these lines might suggest. For one thing, there is still an on-going debate on where exactly the line between semantics and pragmatics has to be drawn. For instance, dynamic semantics, that was developed in the 80-ties, shifts some issues traditionally belonging to pragmatics back into semantics. Furthermore, for many concrete observations on the meaning of natural language expressions it is still unclear whether they should be explained as semantic or pragmatic phenomena. In this dissertation we will discuss three such observations on the interpretation of English sentences for which the question how to account for them, in particular, whether they are effects of semantic or pragmatic meaning, is still conceived as open. For all of them we will develop a theory taking a very specific standpoint with respect to the semantics-pragmatics distinction. Although some of the general ideas underlying these theories are not new, the work presented here differs from other approaches that follow similar lines in the grade of elaboration of these ideas.

The first observation we want to account for is the *Free choice inference* of disjunctive modal sentences. It has often been observed that sentences like (1) allow the hearer to conclude that both taking a pear and taking an apple are permissible options.

- (1) You may take an apple or a pear.

Standard semantic theories have problems in accounting for this observation. We will develop the idea that free choice inferences are actually pragmatic inferences. More particularly, we will account for them as conversational implicatures. One of the major criticisms Grice's theory of conversational implicatures has to face is that it is not able to make precise predictions. We will therefore first develop a partial formalization of this theory, and then show that this formalization allows us to account for the free choice inferences.

The second phenomenon that we will discuss is the particular way we often enrich what is standardly assumed to be the semantic meaning of answers. For illustration, in a dialogue like (2) Bob's answer is often interpreted as exhausting the

predicate in question, hence, as stating not only that John and Mary passed the examination, but also that these are the only people that did.

(2) Ann: Who passed the examination?

Bob: John and Mary.

This reading is called the *exhaustive interpretation* of answers. In Chapter 3 of this dissertation a formal description of this phenomenon is developed that respects its non-standard logical properties and also accounts for dependencies on the form of the answer given and the contextual relevance of the answer. We will argue that the exhaustive interpretation of English answers is part of the pragmatic meaning of answers, more particularly, a conversational implicature. We will support this claim by proving that a simplified version of the description of exhaustive interpretation provided can be derived from the formalization of conversational implicatures introduced in Chapter 2.

Finally, we will discuss the meaning of English conditional sentences. The aspect of the meaning of these constructions that interests us here are primarily their temporal properties. More particularly, we want to explain the apparent discrepancies between the form of English conditional sentences – especially the tense morphology occurring in them – and the temporal interpretation they obtain. For instance, in so-called subjunctive conditionals like (3) the antecedent is marked with the simple past. However, the antecedent cannot be interpreted as referring to the past.

(3) If you asked him, Peter would help you.

We will develop an approach that derives these temporal properties compositionally from the meaning of the parts of the construction. Thus, in contrast to the first two topics, in this case we will make semantics responsible for the observations under debate.<sup>1</sup>

But before we start to consider the temporal properties of English conditional sentences, we will first, in Chapter 5, discuss the meaning of these sentences on a more abstract level that ignores time. The reason is that there are some open questions concerning the meaning of in particular counterfactual conditionals that have to be answered before we can properly account for the temporal properties of conditionals. After this has been done we will, in Chapter 6, extend the timeless framework developed in Chapter 5 with (i) the introduction of a more complex logical form for conditionals with formal expressions for the English tenses, the

---

<sup>1</sup>I do not intend to claim that there are no pragmatic aspects to the meaning of conditionals in general, not even that all their temporal properties can be explained purely based on semantics. The claim rather is that the temporal properties *under discussion* are due to the semantic part of the meaning of English conditional sentences.

perfect, and modals *will*, *would*, *may* and *might*, and (ii) the addition of time to the model with respect to which the logical form is interpreted. We will provide a compositional semantics for this logical form that correctly accounts for the temporal properties of conditionals under discussion.

Besides their relevance for the semantics-pragmatics debate, there is another way in which the three topics discussed in this book are connected. In all three cases the interpretation of sentences will be described using *minimal models*. Let us be a bit more explicit on what we mean with the use of minimal models. Assume that you have defined a function  $I$  that assigns interpretations to sentences  $\psi$  of some formal language  $\mathcal{L}$ . More precisely, the function  $I$  is proposed to map elements of  $\mathcal{L}$  on subsets of some domain  $M$ , which is a class of models for  $\mathcal{L}$ -sentences.<sup>2</sup> Then we can strengthen the interpretation function  $I$  by defining a new interpretation function  $I^*$  that maps a sentence  $\psi$  of  $\mathcal{L}$  to some subset of  $I(\psi)$ . This subset can be defined, for instance, as the set of minimal elements of  $I(\psi)$  with respect to some order  $\leq$  on  $M$ :  $I^*(\psi) = \text{Min}(\leq, I(\psi))$ .<sup>3</sup> Such a strengthening of a basic interpretation function  $I$  by selecting minimal models will stand central in our account for all three phenomena discussed in this book: free choice inferences, exhaustive interpretation, and conditionals.

The use of minimal models has been introduced in Artificial Intelligence to model certain non-monotonic aspects of practical reasoning. The observation that was to be captured is that in every-day reasoning we tend to jump to conclusions that are not warranted by classical deductive logic. This reasoning strategy can be described as selecting only a subclass of the standard, deductive models for a set of premises. The idea driving the minimal model approach is that the relevant subclass of models are those that are minimal in some respect. For instance, the formalism of *predicate circumscription*, as special instantiation of the minimal models approach, selects models that assign minimal extensions to certain relevant predicates. Because of its intuitive cognitive plausibility, minimal models are still a popular approach to non-monotonic reasoning. In semantics it is, as non-monotonic reasoning techniques in general, less well-known. However, minimal models are standardly used in the description of the meaning of conditional sentences and have been introduced for this application years before they were ‘re-invented’ by researchers in Artificial Intelligence (see, for instance, McCarthy 1980, 1986).

In the second and the third chapter we will use minimal models primary to formalize pragmatic reasoning. More particularly, we will use them to make

---

<sup>2</sup>In the following we will call  $M$  or the structure  $M$  is part of a *model* for some formal language and the elements of  $M$  *possible worlds* or *possibilities*. We use here and in the title the term *minimal models* instead of *minimal worlds* or *minimal possibilities* because it is the less technical and more intuitive notion.

<sup>3</sup>In this formula  $\text{Min}$  denotes the operation that selects minimal elements in the set  $I(\psi)$  with respect to the order  $\leq$ .

parts of Grice's theory of conversational implicatures concrete. In this context the function  $I$  will refer to semantic meaning of a sentence and  $I^*$  to a strengthening of semantic meaning with pragmatic information. In the second part of the book, the Chapters 4, 5, and 6, minimal models will be used to model the semantic meaning of conditional sentences. As standard in the literature, we will claim that a conditional with antecedent  $A$  and consequent  $C$  is true in a world  $w$ , if the consequent holds on those worlds making the antecedent true that are most similar to  $w$ . These most similar worlds are defined as the minimal models with respect to some order comparing similarity. Also in this context, the function  $I$  refers to an (abstract version) of semantic meaning. But  $I^*$  is a semantic interpretation function as well. The operation  $*$  is proposed to be part of the meaning of the conditional connective. A central contribution of the present work on conditionals lies in the way it specifies the similarity relation – and thereby the operation  $*$ . We claim that laws, in particular causal laws, play an important role for similarity.

**Editorial remarks.** Before we can start with a detailed discussion of the three topics, some final comments on the form of the thesis are in order. The parts of this book that deal with the free choice inferences and exhaustive interpretation have been already published, Chapter 2 on free choice as Schulz (2005) and Chapter 3 on exhaustivity as Schulz & van Rooij (2006). The articles are reprinted here with the kind permission from Springer Science and Business Media. The material presented in the Chapters 4, 5 and 6 on the meaning of English conditional sentences has not been published before.

The article reprinted in Chapter 3 presents joint work with Robert van Rooij. According to the promotion regulations of the University of Amsterdam I have to clarify in this case the contributions made by each of the authors. This is not easily done. The paper emerged from close cooperation and represents the result of extensive discussions between both authors. For some central claims we can at least reconstruct where the basic ideas came from. The observations on the relevance dependence of exhaustive interpretation together with the provided formalization using decision theory origins in work of Robert van Rooij. The basic ideas of the provided formalization of Grice's theory of conversational implicatures are due to the author (see Chapter 2), as is the result on the relation between the proposed formal description of exhaustive interpretation and this formalization of Grice.





## Chapter 2

---

# A pragmatic solution for the paradox of free choice permission

## 2.1 Introduction

(4) *You may go to the beach or go to the cinema*

I almost told my son Michael. But I thought better of it, and said:

(5) *You may go to the beach.*

Boys shouldn't spend their afternoons in the stuffy dark of a cinema, especially not with such lovely weather as to-day's. Thus, what I did in fact permit was less than what I first intended to permit. We might even be inclined to say that the permission I contemplated, entailed, but was not entailed by, the permission I gave. [Kamp (1973), p. 57]

These are the starting lines of a paper of Kamp from 1973 with which he illustrated the well-known phenomenon of *free choice permission*: a sentence of the form *You may A or B* seems to entail the sentences *You may A* and *You may B*.<sup>1</sup>

According to the logical paradigm, a theory of interpretation should provide a formal description of the intuitive inferences a sentence of English comes with, thus, as we will say, it should lay down the *logic* of English.<sup>2</sup> As the extensive

---

<sup>1</sup>This chapter has been published as 'A pragmatic solution for the paradox of free choice permission' 2005 in *Synthese*, 147(2): 343-377. The article is reprinted here with the kind permission from Springer Science and Business Media.

<sup>2</sup>In this chapter we mean by the logic of a language a formally defined notion of entailment between the sentences of the language. The exact form of the definition is unspecified: it may be in terms of a proof system or a model-theoretic description.

literature on the subject shows the inference of free choice permission poses a serious problem for this approach to interpretation. In fact, some students of the problem have argued that it is impossible to come up with a logic of English that treats free choice permission as valid.

Let us take a closer look at one of the central arguments brought forward to support this claim. One way to approach the logic of sentences like (4) and (5) is to describe the meaning of the involved expressions as *may* and *or* by providing an axiomatization of the truth-maintaining reasoning with sentences containing them. However, it seems impossible to find a reasonable set of axioms and derivation rules such that free choice permission becomes a valid inference. As soon as one arrives at a system that together with other necessary and uncontroversial assumptions takes free choice permission to be valid, a range of unintuitive conclusions become derivable as well. For instance, the derivation rules of modus ponens and necessitation, together with the classical tautologies and taking deontic *may* and *must* to be interdefinable<sup>3</sup> seem to be very uncontroversial assumptions. But if the rule of free choice permission is added to this system it allows the following absurd argument (see Zimmermann (2000)).<sup>4</sup>

- (6) a. Detectives may go by bus.
- b. Anyone who goes by bus goes by bus or boat.
- c. Thus, detectives may go by bus or boat.
- d. We conclude that detectives may go by boat.

The apparently unbridgeable misfit between what the logic of sentences like (4), (5), and those in (6) is supposed to look like and the intuitive validity of free choice permission has led von Wright (1969) to speak of a *paradox* of free choice permission. But now one might continue, if there is no convincing logic of English that captures the validity of free choice permission, then the formal approach is not an adequate strategy to describe the semantics of English. Consequently, we should better dismiss the logical paradigm.

At least two assumptions involved in this line of argumentation have been found deficient. First, one can question whether the ‘necessary and uncontroversial’ assumptions about valid semantic inferences of English involved in the argument (6) are actually that uncontroversial. For instance, Zimmermann (2000) has argued that  $A \rightarrow (A \text{ or } B)$  is not valid for the semantics of English, thus, that English *or* cannot be translated as inclusive disjunction  $\vee$ . As a consequence, in

---

<sup>3</sup>In the sense that *You may A* means the same as *It is not the case that you must not A*.

<sup>4</sup>The step from (6a) to (6c) is admissible because one can prove in such a system that from  $A \rightarrow B$  it follows *may A*  $\rightarrow$  *may B*. (6d) is obtained from (6c) by an application of free choice permission.

the example above the step from (6a) to (6c) is not admissible and the implausible conclusion (6d) can no longer be derived.

A different kind of explanation for paradoxes similar to the paradox of free choice permission has been proposed by Grice (1957). He addresses generally the observation that classical logic does not seem to be able to describe the way we interpret English sentences. Grice admits that this is the case. However, he claims, this does not mean that it is not the appropriate logic to model the *semantics* of English. His point is that semantic meaning does not exhaust interpretation. There is also a contribution of contextual *use* to meaning. This information, the *pragmatic* meaning, then closes the gap between the classical logic of semantics and our intuitive understanding of English. Applied to the paradox of free choice permission this means that an axiomatization of the semantics of sentences like (4), (5), and (6) as proposed by von Wright is on the right track. The fact that this logic is incompatible with free choice permission only suggests that this inference should better be analyzed as a pragmatic phenomenon. Grice's plan was then to provide a pragmatic theory that rescues the simple logical approach to language. This enterprise became known as the *Gricean Program*. Grice also outlined parts of such a pragmatic theory in his theory of conversational implicatures. According to this theory a speaker can derive additional information from taking the speaker to behave rationally and cooperatively in conversation. For Grice this means that the speaker will obey certain principles that govern such behavior: the *maxims of conversation*.

So far we have sketched two possible ways out of the paradox of free choice permission: first we can say that the notion of entailment on which the derivation of (6d) from (6a) is based is not the entailment of the semantics of English. Then, of course, we have to provide a better candidate that does not produce such infelicitous predictions. The second option is to follow the Gricean program: we keep the classical logical semantic analysis and propose free choice permission to be a pragmatic phenomenon. Then we are required to come up with a pragmatic theory that can account for the free choice inference. In this chapter we want to explore the second option. This choice has not been adopted based on an evaluation of free choice permission as pragmatic inference. While we will see that many characteristics of this inference speak for such an approach, observations pointing in the opposite direction can be found as well. The theoretical question driving the research was rather whether a satisfying pragmatic explanation for free choice permission *can* be given. There is a well-known and dreaded obstacle such an approach has to overcome. To show that a certain inference can be explained by Grice's theory of conversational implicatures, we first need a precise description of the conversational implicatures an utterance comes with. Grice himself did not provide such a tool. One of the main goals of pragmatics in the last decades has been to overcome this deficiency (e.g. Horn (1972), Gazdar (1979), Hirschberg (1985)), but a completely satisfying proposal in this direction is still missing. One

may ask for the reason of this lack of success. Perhaps Grice's program to rescue the logical approach to semantics only has shifted the problem to the realm of pragmatics. Now it is this part of interpretation that resists a formalization.

There are good reasons to believe that the mentioned attempts to improve on the clarity of Grice's theory did not exhaust their possibilities. When looking at the proposals made it emerges that a rather limited set of technical tools has been used. The main role is still played by classical deductive logic; the logic of Frege and Tarski. But also logic has had its revolutions since their times, among them the development of non-monotonic reasoning. Non-monotonicity has always been considered to be a central feature of conversational implicatures.<sup>5</sup> This suggests that techniques developed in non-monotonic logic may be of use to formalize the theory of Grice. In this chapter we will try to use non-monotonic logic to (partially) formalize Grice's theory of conversational implicatures – at least to the extent that it allows us to give a pragmatic, Gricean explanation of the free choice permission.

Let us summarize the discussion so far. The aim of the present chapter is to provide an explanation of the phenomenon of free choice permission. By *explanation* we mean to come up with a formally precise and conceptually satisfying description of the semantic and pragmatic meaning of expressions like (4) and (5) such that we can explain why the second sentence follows from the first. In the framework of this chapter we are not looking for *any* kind of explanation. The idea is to see how far we can get with a pragmatic explanation along the lines of the Gricean program. Thus, we want to maintain a simple approach to semantics that is based on classical logic. In particular, we will interpret utterance as in (4) and (5) as assertions, *or* as inclusive disjunction, and *may* as a unary modal operator. On the basis of such a semantics free choice permission will not come out as valid. Instead, this inference is to be explained as a conversational implicature. To overcome the lack of precision in the theory of Grice we will try to formalize parts of it using non-monotonic logic. Hopefully, this can be done in a way such that we can account for free choice permission.

The rest of the chapter is structured as follows. In the following section we study in some more detail the phenomenon of free choice permission to get a clearer impression of what we have to explain. Afterwards a new Gricean approach to free choice permission is developed. Then we will discuss the proposal and compare it to other recent accounts of free choice permission. The chapter will finish with conclusions and an outlook on future work.

---

<sup>5</sup>In the linguistic literature they do not refer to this property as *non-monotonicity* but call it the *cancellability* of conversational implicatures. This term has been also used by Grice himself.

## 2.2 Free choice inferences

In this section we will have a closer look at the linguistic phenomenon we want to account for. The aim is to obtain a clear picture of the properties of free choice permission. We will also provide some linguistic motivation for the kind of approach we have adopted.

Part of the simple approach to the semantics of sentences as (4) adopted here is that we take them to be assertions. There have been doubts about such an analysis. Kamp (1973), for instance, defends a proposal that takes such sentences to be performatives, granting a permission. However, a closer look on the data reveals that we at least additionally need an approach to the free choice reading of (4) that treats the sentence as an assertion.

It seems to be quite clear that the problematic sentences *do* have a reportative reading and that also this reading allows to infer free choice permission. Assume one student asks another about the submission regularities concerning some abstract. The answer she gets is (7).

(7) You may send it by post or by email.

This sentence also allows a free choice reading according to which both ways of submission are admitted. But in this context it is clear that it is not the speaker who is granting the permission. Thus, even if we could solve the paradox of free choice permission for the performative use, the problem would still exist for the assertive reading. A similar point is made by the observation that parallel inferences as free choice permission also exist for other constructions that cannot be analyzed as performatives (the examples stem from Kamp (1979)).

(8) a. We may go to France or stay put next summer. (with the epistemic reading of *may*)

b. I can drop you at the next corner or drive you to the bus stop.

Similar to example (4), (8a) seems to entail *We may go to France* and *We may stay put next summer*. In the same way the use of (8b) allows the hearer to infer *I can drop you at the next corner* and *I can drive you to the bus stop*. Zimmermann has also argued that the inference of (9) that Peter may have taken the beer from the fridge and that Mary may have taken the beer from the fridge should be analyzed as belonging to the same family.

(9) Peter or Marie took the beer from the fridge.

We will call all these inferences *free choice inferences*. Their similar structure suggests to treat them all as due to the same underlying mechanism. But then

nothing of this mechanism should hinge on the possible performative use of (4).

The examples above also illustrate that free choice inferences can come with sentences of quite different forms. This makes it hard to find a semantic explanation of the phenomenon. Semantics would expect some part of the construction of (4) to trigger the free choice permission. But as (8a), (8b), and (9) show, an approach taking the sentence mood, the modal *may*, or modalities in general to be responsible for the inferences is doomed to fail.

Another item that immediately suggests itself as responsible for the free choice readings is the connector *or*. Indeed, many semantic approaches to the problem take this starting point. They propose, for instance, that *or* can function as conjunction, thus, that (4) semantically means, or can mean, (roughly) the same as *You may go to the beach and you may go to the cinema*. One problem for such a proposal is that this conjunctive meaning of *or* does not generalize to arbitrary linguistic contexts. For instance, the sentence (10) does not entail that Mr. X must take a boat and that he must take a taxi.

(10) Mr. X must take a taxi or a boat.

It goes often unnoticed that also (10) comes with free choice inferences. The sentence has an interpretation from which one can conclude that Mr. X still may choose which disjunct of (10) he is going to fulfill, i.e. (10) allows us to infer that Mr. X may take a taxi and that he may take a boat. A similar reading also exists for epistemic *must* (cf. Alonso-Ovalle (2004)).

Another property of the free choice inferences that speaks in favor of a pragmatic approach is the fact that they are cancelable: they disappear in certain contexts.<sup>6</sup> For obvious reasons, context-dependence is difficult to handle for semantic approaches to the free choice inferences. But it is what you would expect when free choice inferences are pragmatic inferences, particularly conversational implicatures.

The first kind of context in which they disappear is the classical cancellation contexts: when they contradict semantic meaning or world knowledge. Consider, for instance, (11).

(11) Peter is in love or I'm a monkey's uncle.

From (11), in contrast to (9), one cannot infer that both sentences combined by *or* are possibly true, and, thus, that the speaker might be a monkey's uncle. Intuitively, it is quite clear why this free choice inference is not admissible: because the (human) speaker cannot be (in the strict sense of the word) a monkey's uncle.

---

<sup>6</sup>Thus, exactly speaking, when we say that a sentence gives rise to free choice inferences we mean that it does so in certain contexts.

There is another class of situations where in particular deontic free choice inferences can be cancelled. These are contexts where it is known that the speaker is not competent on the topic of discourse. This can either be clear from the context or be explicitly said by the speaker, as in (12).

(12) You may take an apple or a pear – but I don’t know which.

This sentence does not convey that the addressee has the choice as to which fruit he picks. Instead, the sentence is interpreted as would be expected if *or* means inclusive disjunction (plus the inference that the speaker takes both, taking an apple and taking a pear, to be possibly permitted; this is conveyed by the continuation *but I don’t know which*). This observation suggests that the competence of the speaker plays an important role in the derivation of free choice permission.<sup>7</sup>

As we have seen in this section, free choice permission is part of a wider class of free choice inferences that can come with quite different linguistic constructions. This form independence of the inferences plus their cancellability gives some linguistic support for the decision to try to come up with a pragmatic explanation for their existence. The goal of the next section is then to provide such a pragmatic approach that can account not only for free choice permission, but for free choice inferences and their properties in general.

## 2.3 The approach

### 2.3.1 Introduction

We come now to the main part of the chapter. In the following, a pragmatic approach to the free choice inferences is developed. Given the intention of the chapter to follow the Gricean program, we will adopt a simple and classical approach to the logic of semantic meaning, in particular, *or* will be interpreted as inclusive disjunction and modal expressions are analyzed as unary modal operators. Because this semantics does not account for the free choice inferences, they have to be described as inferences of the pragmatic meaning of an utterance. We will try to describe them as conversational implicatures.

As pointed out in the introduction, if we want to explain certain inferences as conversational implicatures we first need to formalize the latter notion, i.e. to give a precise description of the conversational implicatures an utterance comes with. In order to do so we will use results from non-monotonic logic, particularly work from Halpern & Moses (1984) recently extended by van der Hoek et al. (1999, 2000).

---

<sup>7</sup>Notice that the epistemic free choice inferences cannot be cancelled in the same way. Adding *but I don’t know which* to a sentence like (9) is intuitively redundant and changes nothing (substantial) about its interpretation.



### 2.3.2 The semantics

Before we can start looking for a pragmatic approach to the free choice inferences we first have to be entirely clear about what our classic approach to the semantics of English can do. Therefore, in this section a precise description of this semantics is given. We will introduce a formal language in which we can express sentences as (4) and (5), at least to that extent that we take to be relevant for the free choice inferences. Then, we will provide a model-theory for this language, and, thereby, a semantic theory for the sentences.

*The Language* The semantics of the sentences giving rise to the free choice inferences is formulated in modal propositional logic. Our formal language  $\mathcal{L}$  is generated from a finite set of propositional atoms  $\mathcal{P} = \{\top, \perp, p, q, r, \dots\}$ , the logical connectives  $\neg, \wedge, \vee$ , and  $\rightarrow$ , and two unary modal operators  $\{\Diamond, \Delta\}$ . The diamond is used to formalize epistemic possibility (thus  $\Diamond p$  stands for *possibly p*). The intended reading of  $\Delta p$  is roughly *p is permitted*. We will use  $\nabla$  to shorten  $\neg\Delta\neg$  and  $\Box$  abbreviates  $\neg\Diamond\neg p$ .  $\Box\phi$  is thus true if the speaker believes  $\phi$ . This gives a very simplified picture of the modalities we can express in English. However, we hope that it will become clear that the approach to the free choice inferences we are going to propose applies as well to more complex modal systems.

We call  $\mathcal{L}^0 \subseteq \mathcal{L}$  the language that contains the modal-free part of  $\mathcal{L}$ , i.e. the language defined by the BNF  $\chi ::= p(p \in \mathcal{P}) \mid \chi \wedge \chi \mid \neg\chi$ .<sup>8</sup> Furthermore, we introduce the following abbreviations for certain  $\mathcal{L}$  sentence-schemes:  $[D]$  for  $\Box\phi \rightarrow \Diamond\phi$ ,  $[4]$  for  $\Box\phi \rightarrow \Box\Box\phi$ , and  $[5]$  for  $\neg\Box\phi \rightarrow \Box\neg\Box\phi$ .

*The Semantics* The model theory we assume for  $\mathcal{L}$  is standard for modal propositional logic. A *frame* for  $\mathcal{L}$  is a triple of a set of worlds  $W$  and two binary relations  $R_\Delta$  and  $R_\Diamond$  over  $W$ . A *model* for  $\mathcal{L}$  is a tuple consisting of a frame for  $\mathcal{L}$  and an interpretation function  $V$  for the non-logical vocabulary of  $\mathcal{L}$ : a function from  $p \in \mathcal{P}$  to characteristic functions over  $W$ . Let  $F = \langle W, R_\Diamond, R_\Delta \rangle$  be a frame for  $\mathcal{L}$  and  $M = \langle F, V \rangle$  a model. For  $w \in W$ ,  $R_\Diamond[w]$  denotes the set  $\{v \in W \mid \langle w, v \rangle \in R_\Diamond\}$  and  $R_\Delta[w]$  the set  $\{v \in W \mid \langle w, v \rangle \in R_\Delta\}$ . We call the tuple  $s = \langle M, w \rangle$  for  $w \in W$  a *state*. *Truth* of a sentence  $\phi$  of  $\mathcal{L}$  with respect to a state  $s$  ( $s \models \phi$ ) is defined along standard lines. We will give here only the definition of truth for a formula  $\Delta\phi$ :  $M, w \models \Delta\phi$  iff<sub>def</sub> there is a  $v \in W$  such that  $v \in R_\Delta[w]$  and  $M, v \models \phi$ . A set of formulas  $\Gamma$  is *satisfiable* in a set  $S$  of states if there is some  $s \in S$  where all elements of  $\Gamma$  are true. A set of formulas  $\Gamma$  *entails* a formula  $\phi$  relative to a class of states  $S$  ( $\Gamma \models_S \psi$  iff<sub>def</sub> for all  $s \in S$ :  $s \models \Gamma$  implies  $s \models \psi$ ). If  $\Gamma = \{\phi\}$ , we write  $\phi \models_S \psi$ . Because we intend the given model theory to describe the semantic meaning of  $\mathcal{L}$ -sentences, formulas entailed by  $\models$  from a sentence  $\phi$  have to be understood as being entailed by the semantic meaning of  $\phi$ .

---

<sup>8</sup>Of course, extra rules for  $\rightarrow$  and  $\vee$  can be suppressed because these logical operators can be defined in terms of  $\wedge$  and  $\neg$ .

Let  $\mathcal{S}$  be the set of states that entail the sentence-schemes [4], [5], and [D]. It follows that  $\mathcal{S}$  is the class of states  $s = \langle M, w \rangle$  that have a locally (i.e. in  $w$ ) transitive, euclidian and non-blind<sup>9</sup> accessibility relation  $R_\Diamond$ .<sup>10</sup> In the following we will consider as domain of interpretation only subsets of  $\mathcal{S}$ . Conceptually, this means that we assume that the speaker has positive and negative introspective power, and we exclude the absurd belief state.<sup>11</sup>

*The Free Choice Inferences* Now we can formulate the different free choice inferences we came across in section 2.2 in terms of the formal language  $\mathcal{L}$ . Let us write  $\phi \models_S \psi$  if  $\psi$  can be inferred from the utterance of  $\phi$  in context  $S$ . Let  $p, q$  be  $\mathcal{L}$ -sentences that do not contain any modal operators, i.e.  $p, q \in \mathcal{L}^0$ . In order to model the free choice inferences, the following rules should be valid for  $\models_S$ . ( $\{A|B\}$  has to be read as *A is the premise or B is the premise.*)

$$(D1) \quad p \vee q \models_S \Diamond p \wedge \Diamond q$$

$$(D2) \quad \{\Diamond(p \vee q) | \Diamond p \vee \Diamond q\} \models_S \Diamond p \wedge \Diamond q$$

$$(D3) \quad \{\Box(p \vee q) | \Box p \vee \Box q\} \models_S \Diamond p \wedge \Diamond q$$

$$(D4) \quad \{\Delta(p \vee q) | \Delta p \vee \Delta q\} \models_S \Delta p \wedge \Delta q$$

$$(D5) \quad \{\nabla(p \vee q) | \nabla p \vee \nabla q\} \models_S \Delta p \wedge \Delta q$$

As pointed out in the last section, however, free choice inferences are cancellable: certain additional information can suppress their derivation. That means that we do not want (D1) to (D5) to hold for all  $S \subseteq \mathcal{S}$ . To take care of the observation that free choice inferences do not occur if inconsistent with other information in the context we should add to (D1) to (D3) *iff  $\Diamond p \wedge \Diamond q$  is satisfiable in  $S$* . Because of the special cancellation behavior of deontic free choice inferences we need for (D4) and (D5) the extended condition *iff  $\Diamond p \wedge \Diamond q$  is satisfiable in  $S$  and the speaker is not known to be incompetent in  $S$* .

We allow for the antecedent of the free choice inferences two different logical forms depending on the scope relation between  $\vee$  and the modal operators. The reason is that we do not see clear evidence that excludes one of the forms either from representing the underlying structure of a sentence like (13a) or from giving rise to the free choice inferences. Notice, for instance, that different authors have argued that sentences as (13b) where *or* has explicitly wide scope over the modal expressions do have free choice readings as well.

<sup>9</sup>A state  $s = \langle M, w \rangle$  is non-blind in  $w$  with respect to  $R_\Diamond$  of  $M$  iff<sub>def</sub>  $R_\Diamond[w] \neq \emptyset$ .

<sup>10</sup>For a proof see Blackburn et al. (2001).

<sup>11</sup>The reader may be surprised by the choice to ask only for the local validity of the schemes [D], [4], and [5]. One reason why we do not demand them to be valid in all points of a model is that in this chapter we will never come in a situation where we will talk about belief embedded under other modalities. Furthermore, later on we will consider restrictions on frames that are only plausible when imposed locally.

- (13) a. You may take an apple or a pear.  
 b. You may take an apple or you may take a pear.

### 2.3.3 Introducing the general ideas

The central task of any approach to the free choice inferences is to find a notion of entailment that can take over the role of  $\models_s$  in (D1) to (D5). Of course, the first candidate that comes to mind is the semantic notion of entailment  $\models$ . However, the free choice inferences would not be a problem if  $\models$  would do. Thus, and as we have observed already, the free choice conclusions of (D1) to (D5) are not valid on the semantic models of the respective premises. Following Grice's program, this means that we have to look for a pragmatic notion of entailment that does the job, i.e. we have to find a pragmatic interpretation function such that the conclusions of (D1) to (D5) are valid on the *pragmatic* models of the premises.

But which semantic models does the pragmatic interpretation function have to select to make the free choice inferences valid? Let us, for example, take the inference (D2). There are three types of states  $s = \langle M, w \rangle$  where sentence  $\Diamond(p \vee q)$  is true qua its semantic meaning. In a first class of states there are worlds accessible from  $w$  where  $p$  is true but no worlds where  $q$  holds. This possibility is represented by  $s_1$  in figure 2.1. A second type of states has  $q$ -worlds accessible from  $w$ , but no  $p$ -worlds; for illustration see  $s_2$ . Finally, it may be the case that for both propositions  $p$  and  $q$  there are worlds in the belief state of the speaker in  $s$  where they are true. This type of states is exemplified by  $s_3$  in figure 2.1. Only on the last type of states is the conclusion of (D2) valid, i.e.  $s_3 \models \Diamond p \wedge \Diamond q$ . Thus, we need the pragmatic interpretation to be a function  $f$  that maps the class of semantic models of  $\Diamond(p \vee q)$  on the set only containing states like  $s_3$ .

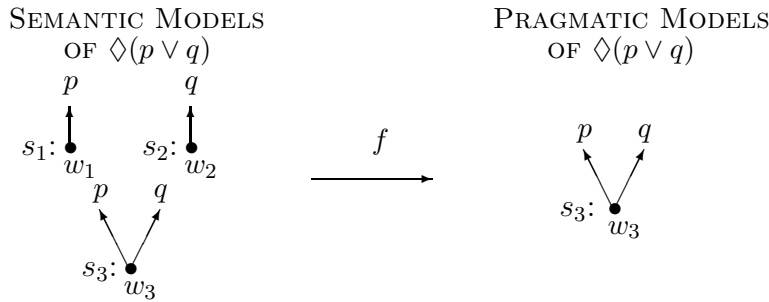


Figure 2.1: The general idea

How can we characterize this function  $f$ ? The central idea of the approach proposed here is that the state  $s_3$  is special because while the speaker believes her utterance to be true she believes less in  $s_3$  than in every other semantic model where this is the case. In  $s_1$ , for instance, the speaker believes more than in  $s_3$

because she holds the additional belief that  $p$  is true. In  $s_2$  compared with  $s_3$  the same holds for  $q$ . Thus, the pragmatic interpretation function  $f$  works as follows: besides  $\Diamond(p \vee q)$  it takes some partial order  $\preceq$  as argument that compares how much the speaker believes in different states, and then it selects those states (i) where  $\Diamond(p \vee q)$  is true qua its semantic meaning, (ii) where the speaker believes her claim  $\Diamond(p \vee q)$  to be true, and (iii) that are minimal with respect to the order  $\preceq$ . More precisely, the pragmatic interpretation  $f_S^\preceq(\phi)$  of a sentence  $\phi$  with respect to a set of states  $S$  and a partial order  $\preceq$  is defined as the set  $\{s \in S \mid s \models \phi \wedge \Box\phi \ \& \ \forall s' \in S : s' \models \phi \wedge \Box\phi \Rightarrow s \preceq s'\}$ . Based on  $f_S^\preceq(\phi)$  we can define the following notion of entailment: we say that sentence  $\phi$  pragmatically entails sentence  $\psi$  with respect to  $S$  and  $\preceq$ ,  $\phi \models_S^\preceq \psi$ , if on all states in  $f_S^\preceq(\phi)$ , i.e. on all pragmatic models of the sentence,  $\psi$  is true.

### 2.3.1. DEFINITION. (The Inference Relation $\models$ )

Let  $\preceq$  be a partial order on some class of states  $S$ . We define for sentences  $\phi, \psi \in \mathcal{L}$ :  $\phi \models_S^\preceq \psi$  iff<sub>def</sub>

$$\forall s \in S : [s \models \phi \wedge \Box\phi \ \& \ \forall s' \in S : s' \models \phi \wedge \Box\phi \Rightarrow s \preceq s'] \Rightarrow s \models \psi.$$

Let us reflect for a moment on the content of this definition. According to  $f$  the interpreter accepts only those models of the speaker's utterance as pragmatically well-formed where the speaker has no additional information that she withholds – by uttering  $\Diamond(p \vee q)$  – from the interpreter. For instance, the interpreter does not take  $s_1$  to be a proper pragmatic model of the sentence. Here, the speaker believes that  $p$  but nevertheless utters the weaker claim  $\Diamond(p \vee q)$ . The interpreter can be understood as taking the speaker to obey the following principle.

*The contribution  $\phi$  of a rational and cooperative speaker encodes all of the information the speaker has; she knows only  $\phi$ .*

Readers familiar with Grice's theory of conversational implicatures will recognize the Gricean character of this assumption. It can be understood as a combination of his maxim of quality with the first sub-clause of the maxim of quantity. To base the free choice inferences on this assumption is to explain them as conversational implicatures. We will therefore call the above statement the *Gricean Principle* and refer to  $f$  as the pragmatic interpretation function *grice*.

### 2.3.4 Working out the details

So far everything has gone quite smoothly. We have localized a Gricean Principle that seems to be responsible for the free choice inferences. We were also able to propose a formalization of the notion of pragmatic entailment this principle gives

rise to. But there is still something missing in definition 2.3.1. We did not define the order  $\preceq$ , i.e. we have not said so far when in some state  $s'$  the speaker believes as least as much as in a state  $s$ . To find a satisfying definition will require some effort.

### 2.3.4.1 The epistemic case

Let us, for a moment, forget about the deontic modalities. When do we want to say that in state  $s' = \langle M', w' \rangle$  the speaker believes as least as much as in state  $s = \langle M, w \rangle$ ? The intuitive answer is that in  $s$  the speaker should be equally or less clear about how the actual world looks like as/than in  $s'$ , thus, she should distinguish in  $s$  a wider range of epistemic possibilities. Or, to be a little bit more precise, every state of affairs she considers possible in  $s$  the speaker should also consider possible in  $s'$ . Then, we have to say what it means that the speaker considers the same state of affairs possible in  $s$  and  $s'$ . Let us try the following: this is the case if there are  $v \in R_\Diamond[w]$  and  $v' \in R_\Diamond[w']$  that interpret the atomic propositions in the same way. Thus we define the order comparing belief states of the speaker as follows.<sup>12</sup>

**2.3.2. DEFINITION.** (The basic order  $\preceq^0$ )

For  $s = \langle M, w \rangle, s' = \langle M'w' \rangle \in \mathcal{S}$  we define  $s \preceq^0 s'$  iff:

$$\forall v' \in R'_\Diamond[w'] \exists v \in R_\Diamond[w] (\forall p \in \mathcal{P} : V(p)(v) = V'(p)(v')).$$

With this definition at hand we can fill out the gap in definition 2.3.1 and obtain the first concrete instance of a pragmatic entailment relation:  $\phi \models_S^{\preceq^0} \psi$ , abbreviated  $\phi \models_S^0 \psi$  holds, if on the  $\preceq^0$ -minimal set of  $S$  where the speaker believes  $\phi$ ,  $\psi$  is valid. This finishes the formalization of the Gricean principle and brings us to the central question of the chapter: can we account for the free choice inferences with this notion of entailment? That means, given that we only consider the epistemic modalities  $\Diamond$  and  $\Box$  in this subsection, are (D1), (D2), and (D3) valid for  $\models_S^0$ ? This is not easily answered. To establish properties of minimal models is not straightforward. The problem is that we have no immediate access to these states. But it turns out that this is not necessary. The only thing we have to show is that any state where the inferences are not valid is not a minimal state.<sup>13</sup>

**2.3.3. FACT.** For any partial order  $\preceq$ , if  $\forall s \in S [s \models \phi \wedge \Box\phi \wedge \neg\psi \Rightarrow (\exists s' \in S : s' \models \phi \wedge \Box\phi \ \& \ s' \prec s)]$ , then  $\phi \models_S^{\preceq} \psi$ .

<sup>12</sup>It is not difficult to prove that the following holds:  $\forall \phi \in \mathcal{L}^0 : s_1 \preceq^0 s_2$  iff  $s_1 \models \phi \Rightarrow s_2 \models \phi$ . Thus the order  $\preceq^0$  could have been defined as well by the condition that in  $s_2$  the speaker believes as least as many  $\mathcal{L}^0$ -sentences as in  $s_1$ .

<sup>13</sup>Fact 2.3.3 holds because the order only compares the belief state for a finite modal depth and we have chosen a finite set of proposition letters. Therefore, we can assume that there are always minimal models.

Fact 2.3.3 tells us that the only thing we have to do to establish, for instance, (D2) (i.e. that for a set  $\{p, q\} \subseteq \mathcal{L}^0$  satisfiable in  $\mathcal{S}$ ,  $\Diamond(p \vee q) \models_S^0 \Diamond p \wedge \Diamond q$  is valid) is to show that for every state  $s \in \mathcal{S}$  that models  $\Diamond(p \vee q) \wedge \Box \Diamond(p \vee q)$  but not the conclusion of (D2) we can find a state  $s^* \in \mathcal{S}$  where still  $\Diamond(p \vee q) \wedge \Box \Diamond(p \vee q)$  is true and  $s^* \prec^0 s$ .

Let  $s = \langle M, w \rangle \in \mathcal{S}$  be a state with the properties described above. Without loss of generality we assume  $s \not\models \Diamond p$ . How can we find the  $s^* \in \mathcal{S}$  we are looking for? This is quite simple: we take  $s^*$  to be the state in  $\mathcal{S}$  that differs from  $s$  only in having an additional world  $\tilde{v}$  in the belief state of the speaker  $R_\Diamond[w^*]$  where  $p$  *does* hold.<sup>14</sup> It is easy to see that this state  $s^*$  has all the properties we need to prove the validity of (D2), i.e. (i)  $s^*$  still models  $\Diamond(p \vee q) \wedge \Box \Diamond(p \vee q)$ , (ii)  $s^*$  is  $\preceq^0$ -smaller than  $s$ :  $s^* \preceq^0 s$ , and (iii)  $s$  is not  $\preceq^0$ -smaller than  $s^*$ :  $s \not\preceq^0 s^*$ .

Ad (i): From  $\langle M^*, \tilde{v} \rangle \models p$  it follows that  $s^* \models \Diamond(p \vee q)$ . Because  $s^* \in \mathcal{S}$  (in particular  $s^* \models [5]$ ) we can conclude that  $s^* \models \Box \Diamond(p \wedge q)$ . Thus  $s^* \models \Diamond(p \vee q) \wedge \Box \Diamond(p \vee q)$ .

Ad (ii): The only difference between  $s^*$  and  $s$  is that  $s^*$  has one more  $\Diamond$ -accessible world:  $\tilde{v}$ . Thus it will clearly be true that  $\forall v \in R_\Diamond[w] \exists v^* \in R_\Diamond[w^*] (\forall p \in \mathcal{P} : V(p)(v) = V(p)(v^*))$ . We can conclude  $s^* \preceq^0 s$ .

Ad (iii): We know that there is a  $v^* \in R^*[w^*]$  such that  $\langle M^*, v^* \rangle \models p$  - this is  $\tilde{v}$ . Because  $s \not\models \Diamond p$  there will be no  $v \in R_\Diamond[w]$  such that  $\langle M, v \rangle \models p$ . Furthermore, because  $p \in \mathcal{L}^0$  in no  $v \in R_\Diamond[w]$  can the interpretation of the atomic propositions be the same as in  $\tilde{v}$ . But that means that  $\forall v^* \in R_\Diamond[w^*] \exists v \in R_\Diamond[w] (\forall p \in \mathcal{P} : V(p)(v) = V(p)(v^*))$  cannot be true. Thus,  $s \not\preceq^0 s^*$ .

Using the same strategy we can also prove that for  $p, q \in \mathcal{L}^0$  such that  $\{p, q\}$  is satisfiable in  $\mathcal{S}$  (D1):  $p \vee q \models_S^0 \Diamond p \wedge \Diamond q$  and  $\Box(p \vee q) \models \Diamond p \wedge \Diamond q$  are valid. But what about the second antecedent of (D3)? Does  $\Box p \vee \Box q \models_S^0 \Diamond p \wedge \Diamond q$  hold? Indeed, it does. Actually, we obtain  $\Box p \vee \Box q \models_S^0 \perp$ . The reason is that there is no  $s \in \mathcal{S}$  such that  $s \models (\Box p \vee \Box q) \wedge \Box(\Box p \vee \Box q)$  and  $s$  is  $\preceq^0$  smaller or equal to every other state in  $\mathcal{S}$  with this property. Thus, we predict that the sentence has no pragmatic models,  $grice_S^0(\Box p \vee \Box q)$  is empty.

To see that there can be no elements in  $grice_S^0(\Box p \vee \Box q)$  notice that  $\phi := (\Box p \vee \Box q) \wedge \Box(\Box p \vee \Box q)$  is, for instance, true in a state where the speaker believes

<sup>14</sup>In Schulz (2004) a constructive description of  $s^*$  is given.  $s^*$  is ‘obtained’ from  $s$  by first adding a world to the model where  $p$  is true – this is possible if  $p$  is satisfiable in  $\mathcal{S}$  – then making this world  $\Diamond$ -accessible from  $w$ , and, finally, close the accessibility relation  $R_\Diamond$  under the axioms [4], [5], and [D] that characterize  $\mathcal{S}$  such that the speaker again gains full introspective power. This closure is important because the state obtained by simply making an additional world  $\Diamond$ -accessible from  $w$  is not an element of  $\mathcal{S}$ . (This also shows that in a strict sense  $s^*$  does not ‘only’ differ in what is  $\Diamond$ -accessible from  $w$ .)

that  $p$  and not  $q$ . Let  $s_1 = \langle M_1, w_1 \rangle$  be a state where this is the case, i.e.  $s_1 \models \Box(p \wedge \neg q)$ . But the sentence is also true if the speaker believes that  $q$  and not  $p$ . Assume that this holds in  $s_2 = \langle M_2, w_2 \rangle$ , i.e.  $s_2 \models \Box(\neg p \wedge q)$ . It is not difficult to see that for  $s_1$  and  $s_2$  neither  $s_1 \preceq^0 s_2$  nor  $s_2 \preceq^0 s_1$  holds. If it were the case that  $\text{grice}(\phi) \neq \emptyset$  (i.e. there would exist a state  $s \in \mathcal{S}$  that models  $\phi$  and for all other states  $s' \in \mathcal{S}$  with this property:  $s \preceq^0 s'$ ) then it would follow that  $s \preceq^0 s_1$  and  $s \preceq^0 s_2$ . By the choice of  $s_1$  ( $s_1 \models \Box \neg q$ ) there are worlds in  $R_\Diamond[w_1]$  where  $q$  does not hold. Because  $p \in \mathcal{L}^0$ , if  $s \preceq s_1$ , i.e.  $\forall v_1 \in R_{1,\Diamond}[w_1] \exists v \in R_\Diamond[w] (\forall p \in \mathcal{P} : V(p)(v) = V_1(p)(v_1))$ , in  $R_\Diamond[w]$  there have to be such worlds too. Thus  $s \models \neg \Box q$ . For the same reason, if  $s \preceq^0 s_2$  there have to be worlds in  $R_\Diamond[w]$  where  $p$  is false, and, hence,  $s \models \neg \Box p$ . But then  $s \models \neg \Box p \wedge \neg \Box q$ . This contradicts the condition  $s \models \Box p \vee \Box q$ . Thus  $\text{grice}(\Box p \vee \Box q) = \emptyset$ .

Conceptually, the fact that for logically independent  $p, q \in \mathcal{L}^0 : \Box p \vee \Box q \models^0 \perp$  means that our theory predicts this sentence to be pragmatically not well-formed. But this seems to be – given Grice’s theory and our formalization thereof – correct. If for a sentence  $\phi$  satisfiable in  $\mathcal{S}$ ,  $\text{grice}_\mathcal{S}^0 = \emptyset$ , then there are incomparable  $\preceq^0$ -minimal states modeling  $\phi \wedge \Box \phi$ . This means that the speaker believes in minimal belief states for  $\phi \wedge \Box \phi$  different things. Then, the speaker has to have in these minimal belief states beliefs she did not communicate. Thus, it is obvious for the interpreter that she did not obey the Gricean Principle. We follow Halpern & Moses (1984) in calling such sentences *dishonest*.

Dishonest sentences provide an interesting testing condition for the theory of Grice and the formalization thereof proposed here. Grice’s theory predicts that dishonest sentences should be pragmatically out: they cannot be uttered by speakers that obey the Gricean Principle. Furthermore, because it is proposed here that the free choice inferences are conversational implicatures, another prediction that can be tested is that the dishonest sentence  $\Box p \vee \Box q$  should not give rise to free choice inferences. And, indeed, sentences like (14a) and (14b) are reported to not allow a free choice reading. In addition, their use seems to be restricted to particular contexts.<sup>15</sup>

(14) a. ?Mr. X must be in Amsterdam or Mr. X must be in Frankfurt.

b. ?I believe that A or I believe that B.

---

<sup>15</sup>One context in which a sentence like (14b) intuitively can be used is when the speaker is known to withhold information and, hence, to be disobeying the Gricean Principle. This is exactly what is predicted by our approach. The following example has been provided by one of the referees.

(i) I know perfectly well what I believe, but all I will say is this: I believe that A or I believe that B.

### 2.3.4.2 The deontic case

As we have seen in the last section we can formalize the Gricean Principle in a way such that we can account for the epistemic free choice inferences in context  $\mathcal{S}$ . But it is easy to see that  $\models_{\mathcal{S}}^0$  will not predict (D4) and (D5) to be valid as well. The reason is that the order  $\preceq^0$  on which this notion of entailment is based and that is intended to compare the beliefs of the speaker does not compare what the speaker believes about the deontic accessibility relation. We said that we want to base the pragmatic interpretation on an order that calls a state  $s \in \mathcal{S}$  smaller than a state  $s' \in \mathcal{S}$  if in the first the speaker believes less/considers more possible than in the second. For the basic information order  $\preceq^0$  (see definition 2.3.2) the only thing that matters is that in the first state the speaker considers more interpretations of the propositional atoms possible than in the second. As a consequence,  $\preceq^0$  compares only the speaker's belief about the interpretation of these atoms (and Boolean combinations thereof).<sup>16</sup> This suggest that to account for the deontic free choice inferences we should extend the order such that it respects also the speaker's beliefs about what holds on the deontic accessibility relation. Thus, we should rather say that in state  $s = \langle M, w \rangle$  the speaker believes less (or equally much) than in state  $s' = \langle M', w' \rangle$  if for every world the speaker considers possible in  $s$  there is some world the speaker considers possible in  $s'$  that not only agree on the interpretation of the propositional atoms but also on which interpretations are deontically possible. This is expressed in the definition of the following order.

#### 2.3.4. DEFINITION. (The Objective Information Order $\preceq^n$ )<sup>17</sup>

For  $s = \langle M, w \rangle, s' = \langle M', w' \rangle \in \mathcal{S}$  we define  $s \preceq^n s'$  iff<sub>def</sub>

$$\begin{aligned} & \forall v' \in R'_{\Diamond}[w'] \exists v \in R_{\Diamond}[w] : \\ & \quad (i) \quad \forall p \in \mathcal{P} : V(p)(v) = V'(p)(v') \text{ \& } \\ & \quad (ii) \quad \forall u \in R_{\Delta}[v] \exists u' \in R'_{\Delta}[v'] (\forall p \in \mathcal{P} : V(p)(u) = V'(p)(u')) \text{ \& } \\ & \quad (iii) \quad \forall u' \in R'_{\Delta}[v'] \exists u \in R_{\Delta}[v] (\forall p \in \mathcal{P} : V(p)(u) = V'(p)(u')). \end{aligned}$$

By substituting  $\preceq^n$  as order in definition 2.3.1 we obtain a new notion of entailment  $\models_{\mathcal{S}}^{\preceq^n}$ , shortly  $\models_{\mathcal{S}}^n$ . In the same way as in the last section one can show that the free choice inferences (D1), (D2), and (D3) are valid for  $\models_{\mathcal{S}}^n$ . The only difference between the orders  $\preceq^0$  and  $\preceq^n$  lays in the conditions (ii) and (iii) which concerns belief about the deontic options. Therefore, they make exactly the same predictions for sentences that do not contain  $\Delta$  or  $\nabla$ .

<sup>16</sup>Actually, this order also respects the speaker's beliefs about the  $\mathcal{L}^0$ -facts. This is due to the fact that in  $\mathcal{S}$  the speaker has full introspective power.

<sup>17</sup> $\preceq^n$  compares only deontic information about basic facts. The order can easily be extended such that it respects all deontic information by using (restricted) bi-simulation (see Schulz (2004)). The reason why we did not choose this more general definition here is that we do not need this complexity. We consider only sentences having in the scope of  $\Delta/\nabla$  a modal free formula.



However, the deontic free choice inferences (D4) and (D5) do not hold for  $\models_{\mathcal{S}}^n$ . Given that (D2) and (D4) show a highly similar structure one may wonder why we can account with  $\models_{\mathcal{S}}^n$  for one but not for the other. The reason is this. In  $\mathcal{S}$  there is no connection between the actual deontic options and the speaker's beliefs about what is deontically accessible. Therefore, from minimizing the speaker's belief the interpreter will learn nothing about what is actually permitted and what not. But the deontic free choice inference  $\Delta p \wedge \Delta q$  is about valid permissions. For the actual epistemic options and the speaker's beliefs about them such a connection is built into  $\mathcal{S}$ . We defined  $\mathcal{S}$  as those states where the speaker has full introspective power. Thus, we assumed that the speaker knows about her beliefs and her uncertainty. This suggests that to make the deontic free choice inferences valid we would need something similar there too, i.e. the speaker has to know about the valid obligations and permissions. The speaker has to be *competent* on the deontic options.

This conclusion is also supported by an observations we made in section 2.2. There, we have seen that the deontic free choice inferences are cancelled if it is known that the speaker is not competent on the deontic options. Thus, it seems that these inferences really depend on additional knowledge about the competence of the speaker.

### 2.3.4.3 Competence

The considerations at the end of the last section suggest that an additional assumption of the speaker's competence may be the missing link to obtain the deontic free choice inferences. For the formalization of this idea we will rely on Zimmermann (2000). He builds on a proposal of Groenendijk & Stokhof (1984) and defines competence by the following first-order model condition.<sup>18</sup>

#### 2.3.5. DEFINITION. (Competence)

A speaker is *competent* in a state  $\langle M, w \rangle \in \mathcal{S}$  with respect to a modality  $\Delta$  iff<sub>def</sub>

$$\forall v \in W^M [v \in R_{\Diamond}^M[w] \Rightarrow (R_{\Delta}^M[v] = R_{\Delta}^M[w])].$$

It is easy to prove that this condition is characterized in modal propositional logic by the two axioms  $[C_1]$ :  $\nabla \phi \rightarrow \Box \nabla \phi$  and  $[C_2]$ :  $\neg \nabla \phi \rightarrow \Box \neg \nabla \phi$ , i.e. a speaker is competent in some state  $s = \langle M, w \rangle$  if the underlying frame locally (hence, in  $w$ ) satisfies  $[C_1]$  and  $[C_2]$ .  $[C_1]$  is a generalization of axiom [4] formalizing positive introspective power to the multi-modality case; it warrants that the speaker knows about all valid obligations.  $[C_2]$ , on the other hand, generalizes axiom [5] formalizing negative introspective power; it assures that the speaker also knows about the valid permissions.

---

<sup>18</sup>The (intensional) predicate  $\lambda w \lambda x. P(w)(x)$  in his definition is instantiated here by the characteristic function of  $\Delta$ -accessible worlds  $\lambda w \lambda v. w R_{\Delta} v$ .

Let us call  $\mathcal{C}$  the set of states where additionally to the axioms  $[D]$ ,  $[4]$ , and  $[5]$  also the competence axioms  $[C_1]$  and  $[C_2]$  are valid. Do we get the free choice inferences for  $\models_{\mathcal{C}}^n$ ? Unfortunately, this is not the case. The pragmatic interpretation we obtain this way is much too strong. It is predicted that every sentence  $\phi \in \mathcal{L}$  satisfiable in  $\mathcal{C}$  gives rise to an empty pragmatic interpretation, i.e. is dishonest. Or, in other words, given the way  $\models_{\mathcal{C}}^n$  interprets the Gricean Principle a speaker competent on  $\Delta$  as formalized in  $[C_1]$  and  $[C_2]$  cannot utter any non-absurd sentence and be obeying this principle.

Let us have a closer look at why this is the case. Given the formalization of competence we have chosen, a competent speaker knows for every  $\chi \in \mathcal{L}$  which of the sentences  $\nabla\chi$  and  $\neg\nabla\chi$  holds. Hence, in all states of  $\mathcal{C}$  and for all sentences  $\chi \in \mathcal{L}$  either  $\Box\nabla\chi$  or  $\Box\neg\nabla\chi$  is true. However, it is easy to see that for every  $\chi \in \mathcal{L}^0$  a state where  $\Box\nabla\chi$  holds is  $\preceq^n$ -incomparable with a state where  $\Box\neg\nabla\chi$  holds. Thus, to prevent dishonesty, i.e. to warrant that the interpreter does not end up with different incomparable minimal states, for the sentence  $\phi$  uttered by the speaker either  $\phi \wedge \Box\phi \models_{\mathcal{C}} \Box\nabla\chi$  or  $\phi \wedge \Box\phi \models_{\mathcal{C}} \Box\neg\nabla\chi$  has to hold. But the same argument applies for every  $\chi \in \mathcal{L}^0$ ! Thus, for every sentence  $\chi \in \mathcal{L}^0$  it has to be the case that  $\phi$  entails semantically either that the speaker believes  $\nabla\chi$  or that she believes  $\neg\nabla\chi$ . There can be no finite and satisfiable sentence that is that strong. Hence, every sentence  $\phi \in \mathcal{L}$  satisfiable in  $\mathcal{C}$  is dishonest.

#### 2.3.4.4 Solving the paradox of free choice permission

One way to look at the problem we ended up with in the last section is that the formalization of the Gricean Principle given with  $\models^n$  is too strong. By  $\models^n$  a speaker who wants to obey the principle has to give every bit of information about deontic accessible interpretations of the basic atoms that she has. Perhaps we can obtain a more natural notion of pragmatic entailment when we allow the speaker to withhold some of this information. The problem, then, becomes to find the right restriction that fits our intuitions.

To start with, we can ask ourselves which information about the deontic accessibility relation we can take to be not relevant for the order because it is accessible to the interpreter anyway. It turns out that if the speaker is competent on  $\Delta$ , then which permissions the speaker believes to hold can be already concluded from taking her to convey all she knows about the valid obligations. If she is honest about this part of her beliefs, then if her utterance  $\phi$  does not entail for some  $\chi \in \mathcal{L}^0$  that she believes  $\nabla\chi$  she cannot believe this obligation to be valid, i.e.  $\neg\Box\nabla\chi$  holds. From her competence it follows that she has to believe that  $\neg\chi$  is permitted. On the other hand, if for some  $\chi \in \mathcal{L}$  it holds that the speaker believes  $\chi$  to be permitted, then, by competence,  $\neg\Box\nabla\neg\chi$  is true and because we assume her to believe in her utterance  $\phi$ ,  $\phi$  cannot entail  $\nabla\neg\chi$ . Thus, a competent speaker believes some sentence  $\chi \in \mathcal{L}^0$  to be permitted if and only if her utterance does not entail that  $\chi$  is prohibited. This suggests that information

about which permissions the speaker believes to be valid can be ignored by the order. It is enough to compare what a competent speaker believes to be a valid obligation.<sup>19</sup> We obtain such an order when we delete condition (ii) from the definition of  $\preceq^n$ .<sup>20</sup>

**2.3.7. DEFINITION.** (The Positive Information Order  $\preceq^+$ )<sup>21</sup>

For  $s = \langle M, w \rangle, s' = \langle M', w' \rangle \in S$  we define  $s \preceq^+ s'$  iff<sub>def</sub>

- $$\begin{aligned} & \forall v' \in R'_\Diamond[w'] \exists v \in R_\Diamond[w] : \\ & (i) \quad \forall p \in \mathcal{P} : V(p)(v) = V'(p)(v') \text{ \& } \\ & (ii) \quad \forall u' \in R'_\Delta[v'] \exists u \in R_\Delta[v] (\forall p \in \mathcal{P} : V(p)(u) = V(p)(u')). \end{aligned}$$

By substituting  $\preceq^+$  in definition 2.3.1 we obtain a new notion of pragmatic entailment:  $\models_s^+$ , abbreviated  $\models_s^+$ . It turns out that for  $\models_s^+$  not only the free choice inferences for the epistemic modality are valid, but (D4) and  $\nabla(p \vee q) \models_c^+ \Diamond\Delta p \wedge \Diamond\Delta q$  as well. Parallel to the epistemic case the sentence  $\nabla p \vee \nabla q$  is predicted to be dishonest when uttered by a competent speaker that obeys the Gricean Principle.

Let us discuss the validity of (D4). The argumentation we employ has exactly the same structure as in section 2.3.4.1. If for  $p, q \in \mathcal{L}^0$  such that  $\{p, q\}$  is satisfiable in  $\mathcal{C}$  (D4):  $\Delta(p \vee q) \models_c^+ \Delta p \wedge \Delta q$  were not valid then there would be a state  $s \in \mathcal{C}$  minimal with respect to  $\preceq^+$  such that  $s \models \Delta(p \vee q) \wedge \Box\Delta(p \vee q)$  but not  $s \models \Delta p \wedge \Delta q$ . Now, we show that this cannot be the case: every state  $s \in \mathcal{C}$  that semantically entails  $\Delta(p \vee q) \wedge \Box\Delta(p \vee q)$  but where the consequence of (D4) is not true cannot be minimal with respect to  $\preceq^+$ .

Assume that for  $s = \langle M, w \rangle \in \mathcal{C}$  we have  $s \models \Delta(p \vee q) \wedge \Box\Delta(p \vee q)$ , but  $s \not\models \Delta p \wedge \Delta q$ . Without loss of generality  $s \not\models \Delta p$ . Let  $s^* = \langle M^*, w^* \rangle \in \mathcal{C}$  be

<sup>19</sup>Of course, the same argument can be also used to show that the speaker does not have to convey all she believes about valid obligations, as long as she is honest about her beliefs concerning permissions. However, minimizing beliefs on permissions does not result in a convincing notion of pragmatic entailment. For instance, this one wrongly predicts that sentences like  $\Delta(p \vee q)$  are dishonest. One would like to have some motivation for the choice of the order  $\preceq^+$  besides the fact that it does the job, while some equally salient alternatives do not – particularly, given that we formalize a theory of rational behavior. But so far I am not aware of any conclusive arguments.

<sup>20</sup>Also for this order an equivalent definition using a set of sentences can be given (for a close discussion see Schulz (2004)).

**2.3.6. FACT.** Let  $\mathcal{L}^+ \subseteq \mathcal{L}$  be language defined by the BNF-form  $\chi_+ ::= p(p \in \mathcal{L}^0) | \chi_+ \wedge \chi_+ | \chi_+ \vee \chi_+ | \nabla p(p \in \mathcal{L}^0)$ . Then we have for  $s, s' \in \mathcal{C}$ :

$$s \preceq^+ s' \Leftrightarrow \forall \chi \in \mathcal{L}^+ : s \models \Box\chi \Rightarrow s' \models \Box\chi.$$

<sup>21</sup>Again,  $\preceq^+$  only compares beliefs about formulas  $\{\nabla\chi | \chi \in \mathcal{L}^0\}$ , but an extension to sentences  $\nabla\chi$  for  $\chi \in \mathcal{L}$  is easily possible (see Schulz (2004)). We use the simpler variant because the sentences we consider here are only of the former type.

the state that is like  $s$  except that from  $w^*$  an additional world  $\tilde{v}$  is  $\Delta$ -accessible where  $p$  is true.<sup>22</sup> Thus  $s^* \models \Delta p$ . We show that (i)  $s^* \models \Delta(p \vee q) \wedge \Box \Delta(p \wedge q)$ , (ii)  $s^* \preceq^+ s$ , and (iii)  $s \not\preceq^+ s^*$ . Then  $s$  cannot be minimal because  $s^*$  is smaller.

Ad (i) We have seen already that  $s^* \models \Delta p$ . It follows  $s^* \models \Delta(p \vee q)$ . Because  $s^*$  is an element of  $\mathcal{C}$  we can conclude from this (by  $[C_2]$ ) that  $s^* \models \Box \Delta(p \vee q)$ . This shows (i).

Ad (ii) We have to show that for all  $v \in R_\Diamond[w]$  we can find a  $v^* \in R_\Diamond^*[w^*]$  such that (i)  $\forall p \in \mathcal{P} (V(p)(v) = V'(p)(v'))$  and (ii)  $\forall u' \in R'_\Delta[v'] \exists u \in R_\Delta[v] (\forall p \in \mathcal{P} : V(p)(u) = V(p)(u'))$ . (i) is simple, let us go directly to the interesting case: (ii). Because the difference between  $s^*$  and  $s$  is that  $s^*$  has one more  $\Delta$ -accessible world:  $\tilde{v}$ , we have  $R_\Delta[w] \subset R_\Delta^*[w^*]$ . From  $s, s^* \in \mathcal{C}$  we conclude  $\forall v \in R_\Diamond[w] : R_\Delta[v] = R_\Delta[w]$  and  $\forall v^* \in R_\Diamond^*[w^*] : R_\Delta^*[v^*] = R_\Delta[w^*]$ . Together, this gives:  $\forall v \in R_\Diamond[w] \forall v^* \in R_\Diamond^*[w^*] : R_\Delta[v] \subset R_\Delta^*[v^*]$ . Because by assumption  $s$  and  $s^*$  do not differ in the interpretation assigned in worlds of  $R_\Delta[v]$  to elements of  $\mathcal{P}$  this proves the claim.

Ad (iii) Finally,  $s \not\preceq^+ s^*$ . Because  $s \not\models \Delta p$  we obtain by  $[C_1]$  that  $s \models \Box \neg \Delta p$ . Hence, for no  $v \in R_\Diamond[w]$  and no  $u \in R_\Delta[v]$  we have  $\langle M, u \rangle \models p$ . But from  $s^* \models \Delta p$  with  $[C_2]$  it follows  $s^* \models \Box \Delta p$ , and, thus,  $\forall v^* \in R_\Diamond^*[w^*] \exists u^* \in R_\Delta^*[v^*] : \langle M^*, u^* \rangle \models p$ . Because  $p \in \mathcal{L}^0$  condition (ii) of the definition of  $\preceq^+$  is violated for  $s \preceq^+ s^*$ .

Thus, we see that adopting  $\models^+$  as a formalization of the Gricean Principle and applying it to the set of states  $\mathcal{C}$  where the speaker is competent accounts for the free choice inferences.<sup>23</sup>

<sup>22</sup>Again, Schulz (2004) provides a formally precise version of this proof, including a constructive description of  $s^*$ .  $s^*$  is obtained from  $s$  by first adding a world to the model where  $p$  is true – this is possible if  $p$  is satisfiable in  $\mathcal{C}$  – then making this world  $\Delta$ -accessible from  $w$ , and, finally, close the resulting accessibility relations  $R'_\Diamond$  and  $R'_\Delta$  under the axioms [4], [5], [D],  $[C_1]$ , and  $[C_2]$  to obtain a state that belongs to  $\mathcal{C}$ .

<sup>23</sup>There is another way to repair  $\models_{\mathcal{C}}^n$  such that one can account for the deontic free choice inferences. Instead of weakening the order and thereby be less strict on what a speaker has to convey with her utterance, we can also take her to be less competent. It turns out that the competence axiom we have to drop is  $[C_2]$ : we weaken  $\mathcal{C}$  to the set of states  $\mathcal{C}^+$  where  $[D]$ , [4], [5], and  $[C_1]$  are valid. In this case, the speaker knows all valid obligations, but she may be not aware of certain permissions. While this accounts for the free choice inferences, other predictions made by  $\models_{\mathcal{C}^+}^n$  are less convincing than what is predicted by  $\models_{\mathcal{C}}^+$ . For a more elaborate discussion the reader is referred to Schulz (2004).

Finally, it is interesting to note, that also the combination of  $\models^+$  with  $\mathcal{C}^+$ , hence, the combination of weakening the order and weakening the notion of entailment allows us to derive the free choice inferences. Also this combination of a concept of competence with a formalization of the Gricean Principle does not work as well as  $\models_{\mathcal{C}}^+$ .

### 2.3.5 The cancellation of free choice inferences

In the last sections we have developed a pragmatic notion of entailment that with respect to the set  $\mathcal{S}$  makes the epistemic free choice inferences (D1) to (D3) valid, and with respect to the more restricted context  $\mathcal{C}$  additionally validates the deontic free choice inferences. Have we, thereby, achieved our initial goal to provide a Gricean account for the free choice inferences? No, there is still something to be done. As discussed in section 2.3.2 the free choice inferences are non-monotonic inferences: they can be cancelled by additional information. It remains to be checked whether the approach developed above predicts (D1) - (D5) to be valid exactly in those contexts where such canceling information is not given.

In section 2.2 we have seen that there are two different types of information that may lead to a suspension of free choice inferences. Let us proceed by discussing both of them separately. Our first observation was that free choice inferences are cancelled in case they are inconsistent with information in the context or given by the speaker.<sup>24</sup> It is easy to see that this is also predicted by the system we propose. If one of the consequents of (D1) to (D5) is inconsistent with information in some context  $S$  or the semantic meaning of the utterance made, then there will be no state where this consequent holds among those states in  $S$  where the utterance is true (by its semantic meaning). In particular, the states selected by our pragmatic interpretation function *grice* will not make such a consequent true. Thus, we see that the approach immediately accounts for this part of the non-monotonicity of the free choice inferences.

Now we come to the second observation. As we have seen in section 2.2, the deontic free choice inferences can also be cancelled by information that the speaker is not fully competent on the topic of discourse. Therefore, we should derive these inferences only in contexts where such information has not been given. Whether the proposal made accounts for this observation is not clear yet. We predict the deontic free choice inferences to be valid in a context where the interpreter takes the speaker to be competent and to obey the Gricean Principle. Of course, information that the speaker is in some respects incompetent stands in conflict with taking the speaker to be competent (as described by  $[C_1]$  and  $[C_2]$ ). But we have not said anything so far about how the interpreter behaves in such a situation.

Let us sketch one position one could adopt. We can propose that taking the speaker to be competent is an assumption interpreters make – just as they assume the speaker to obey the Gricean Principle. Interpreters do not make this assumption if they are facing contradicting information.<sup>25</sup> This proposal predicts

---

<sup>24</sup>This is probably the least disputed property characterizing conversational implicatures. Therefore, insofar as we claim to formalize conversational implicatures, all pragmatic inferences we predict should have this property.

<sup>25</sup>Given that the derivation of the free choice inferences appears to be the normal interpre-

that if an interpreter who does not know the speaker to be competent encounters information contradicting the competence assumption, then she will not derive the deontic free choice inferences. If, however, no such conflicting information is given, the interpreter assumes the speaker to be competent on  $\Delta$  and the inferences become valid. So far the cancellation behavior of the deontic free choice inferences is captured correctly. It may, however, be the case that the interpreter knows that the speaker is competent in some respects and that this information does not contradict what she now learns about the incompetence of the speaker. In such a situation it does not seem to be plausible to take this independent information to be cancelled together with the competence assumption. If it is not dismissed then it depends on what exactly the interpreter knows about competence and incompetence of the speaker whether the deontic free choice inferences are derived. This approach needs to be evaluated by comparing its predictions with the interpretational behavior of native speakers. This has to be investigated in future work.<sup>26</sup>

### 2.3.6 Conclusions

In this section we have developed a formalization of the Gricean Principle that can (given standard assumptions about the introspective power of the speaker) account for the epistemic free choice inferences. However, this formalization on its own is not able to derive the deontic free choice inferences as well. They can be predicted if in the context it is additionally known that the speaker is competent on  $\Delta$ . We adopted a strong notion of competence: the speaker is taken to know the valid obligations as well as as all permissions. With this system we can account for all free choice inferences.

Furthermore, we have seen that the proposal also models correctly the cancellation of free choice inferences when conflicting information is encountered. Whether it can also account for the suspension of the deontic free choice given information that the competence of the speaker is limited depends on how we understand the role of the competence assumption in interpreting utterances. We have sketched one possible position that promises to model the cancellation behavior correctly. Empirical investigations have to show whether this proposal is convincing.

---

tation of sentences like (4) *You may go to the beach or go to the cinema*, this position is much more convincing than proposing that the interpreter *knows* the speaker to be competent when inferring free choice.

<sup>26</sup>There are other ways of how we can understand the role of competence in the derivation of the free choice inferences. In the scenario sketched above we took it to be an extra assumption that is cancelled *completely* if conflicting information is encountered. We might as well propose that in such a situation the interpreter tries to maintain as much of the competence assumption as she can. Such an approach has been adopted – for independent reasons – in van Rooij & Schulz (2004). In this case it depends on the kind of information about the incompetence of the speaker the interpreter has whether the deontic free choice inferences are cancelled or not.

## 2.4 Discussion

In the last section we have seen that based on a classical logical approach to the semantics of English the free choice inferences can be described in a formally precise way as due to taking the speaker (i) to obey the Gricean Principle, and (ii) to be competent on the topic of discourse. Thus, the central goal with which we started the chapter has been reached: we came up with an approach to the free choice inferences on the lines of the Gricean program. In the following section we will address some open questions concerning the introduced approach and relate the proposal to other approaches to the free choice inferences.

### 2.4.1 An open problem

Unfortunately, in the present form the approach predicts, along with the free choice inferences, many inferences that are not welcome. For instance, for arbitrary, in  $\mathcal{S}$  logically independent  $p, q, r \in \mathcal{L}^0$  it holds that  $\Delta(p \vee q) \models_{\mathcal{S}}^+ \Diamond r \wedge \Diamond \neg r \wedge \Diamond \Delta r \wedge \Diamond \Delta \neg r$  and  $\Delta(p \vee q) \models_{\mathcal{C}}^+ \Delta r \wedge \Delta \neg r$ . Or, to use more natural examples, we obtain, for instance, that (15a)  $\models_{\mathcal{C}}^+$ -entails (15b) and (15c). These predictions are certainly wrong.

- (15) a. You may take an apple or a pear.  
       b. You may take a banana.  
       c. Aunt Hetty may be making pie.

Where do these strange predictions come from? The pragmatic interpretation function  $grice^+$  on which  $\models^+$  is based selects among the semantic models of a sentence those where the speaker believes the sentence to hold and has as few as possible other beliefs. This is what the Gricean Principle demands: a speaker does not withhold information – *any* information – she has from the hearer.<sup>27</sup> Therefore, it is not surprising that if a speaker utters a sentence like (15a) that does not exclude that aunt Hetty is making pie, then  $\models^+$  predicts that the speaker considers it as possible that she is: according to the Gricean Principle, if the speaker believed that aunt Hetty is not making apple pie, then she would have shared her belief with the audience. She did not do so when uttering (15a). Thus, she cannot hold this belief. The point is that when we interpret utterances, we certainly do not expect the speaker to convey *all* of her beliefs (that are not commonly known). The Gricean Principle underlying  $\models^+$  is too strong.

There is a way out of this problem already suggested in Grice's formulation of the first sub-clause of the maxim of quantity:<sup>28</sup> 'Make your contribution as

<sup>27</sup>The way we have defined the order  $\preceq^+$  'any information' means any information that can be expressed with the following sentences  $\chi ::= p(p \in \mathcal{L}^0) | \chi \vee \chi | \chi \wedge \chi | \nabla p(p \in \mathcal{L}^0)$ .

<sup>28</sup>Thus, our reformulation of this maxim in the Gricean Principle is not entirely faithful to Grice.

informative as required (for the current purpose of exchange)’ (Grice 1989, p. 26). What the Gricean Principle misses is some restriction to contextually required or *relevant* information. Thus, it should rather be formulated as follows.

*The contribution  $\phi$  of a rational and cooperative speaker encodes all of the relevant information the speaker has; she knows only  $\phi$ .*

This suggests that to overcome the above mispredictions we have to formalize contextual relevance and build it into our pragmatic notion of entailment. Some ideas how this can be done can be found in van Rooij & Schulz (2004). In this paper the formalization of the Gricean Principle proposed here is used to give a pragmatic explanation for the phenomenon of *exhaustive interpretation*. *Exhaustive interpretation* describes the often observed strengthening of the semantic meaning of answers to overt questions.<sup>29</sup> In the context of questions it is quite obvious which information is relevant: information that helps to answer the question. The authors propose a version of the interpretation function *grice* that respects such a notion of relevance. In future work it has to be seen whether this solution can be also applied to the modeling of the free choice inferences proposed here.

## 2.4.2 Comparison

### 2.4.2.1 The approaches of Kamp and Zimmermann

The proposal to the free choice inferences introduced in this chapter is highly inspired by the work of Zimmermann (2000) and Kamp (1979) on this subject, particularly the outline of a pragmatic approach of the latter author. Zimmermann, as well as Kamp, bases the free choice inferences on two premisses. The first ingredient is that from a sentence giving rise to free choice inferences the interpreter learns something about the epistemic state of the speaker. From a sentence *You may take an apple or a pear* she learns, for instance, that the speaker takes both, *You may take an apple* and *You may take a pear* to be possibly true. Sometimes, this already accounts for the free choice observation, as for instance, for examples like (9): *Mary or Peter took the beer from the fridge*. But for free choice permission this is not enough. Further information is necessary and both approaches take this to be due to the assumption that the speaker is competent on the deontic options.

The second part, the reliance on competence of the speaker, has been adopted here. But the way these two proposals accounted for the derivation of the first

---

<sup>29</sup>For instance, in many contexts the answer *John* to a question *Who smokes?* is not only understood as conveying that John is among the smokers – what would be its semantic meaning – but it is additionally inferred that John is the only one who smokes.



part, the epistemic information, has been found deficient. Zimmermann takes the semantics of *or* to be responsible. Among other things this leads to unreasonable predictions when *or* occurs embedded under other logical operators. Kamp derived the relevant assumptions on the belief state of the speaker via Grice's maxim of brevity. This approach is not general enough to extend to all contexts in which free choice inferences are observed.<sup>30</sup> Therefore, in the chapter at hand the relevant epistemic inferences are derived in a different way: as conversational implicatures due to the first sub-clause of the maxim of quantity and the maxim of quality, summarized in the Gricean Principle.

#### 2.4.2.2 Gazdar's approach to clausal implicatures

Already Gazdar (1979) analyzed the epistemic inferences that Peter may have taken the beer and Mary may have taken the beer from (9): *Mary or Peter took the beer from the fridge* as effects of the first subclause of Grice's maxim of quantity. Gazdar distinguishes two classes of implicatures due to this maxim. The first class, *scalar implicatures*, is not relevant for the discussion at hand. The inferences of (9) just mentioned fall in Gazdar's class of *clausal implicatures*. This rises the question how Gazdar's approach to these implicatures relates to the description of the inferences proposed here – and whether a combination with an assumption of competence of the speaker leads to the free choice inferences as well.

Gazdar (1979) describes the following procedure to calculate clausal implicatures. First, he defines the set of *potential* clausal implicatures (pcis) of a compound sentence  $\psi$ . The pcis of  $\psi$  are the sentences  $\chi \in \{\Diamond\phi, \Diamond\neg\phi\}$  where  $\phi$  is a subsentence of  $\psi$  such that  $\psi$  neither entails  $\phi$  nor its negation  $\neg\phi$ .<sup>31</sup> But not all potential clausal implicatures are predicted by Gazdar to become part of the interpretation of an utterance. Gazdar proposes that first they have to pass a strict consistency check: Add to the common ground the assumption that the speaker knows her utterance to be true<sup>32</sup> and a set of potential clausal implicatures that is satisfiable in this context. Only those pcis are predicted to be present that are satisfiable in all contexts that can be reached this way.

Given the similarity between both approaches it should not come as a surprise that the predictions made by Gazdar (1979) are strongly related to the ones we obtained in section 2.3. Gazdar is able to predict all epistemic free choice inferences (D1), (D2), and (D3). With a weaker notion of competence than used in section 2.3 his approach is even able to derive the deontic free choice

<sup>30</sup>For a detailed discussion of these two approaches and their shortcomings see Schulz (2004).

<sup>31</sup>Gazdar adopts a slightly different interpretation of the modal operators as is proposed in section 2.3. He takes  $S4$  to be the logic of the modal operator  $\Diamond$ . This is partly due to the fact that for Gazdar  $\Box$  models knowledge and not belief. Gazdar's definition of pcis contains one further condition, but this one can be ignored for our purposes.

<sup>32</sup>This is Gazdar's formalization of the  $\mathcal{T}$ -implicatures an utterance comes with.

inferences (D4) and (D5) for competent speakers and, thus, to account for free choice permission.<sup>33</sup>

Let us run through the calculations for (D4). Gazdar can account for this inference only based on the antecedent giving the disjunction wide scope over the modality:  $\Delta\phi \vee \Delta\psi$ . For this sentence he predicts the following set of pcis:  $\{\Diamond\phi, \Diamond\psi, \Diamond\Delta\phi, \Diamond\Delta\psi$  and the respective negations $\}$ . If we assume the speaker to be competent, i.e take as context the set  $\mathcal{C}$ , then we will not predict free choice permission. In  $\mathcal{C}$  the pcis  $\Diamond\Delta p$  and  $\Diamond\neg\Delta p$ , as well as  $\Diamond\Delta q$  and  $\Diamond\neg\Delta q$  contradict each other and, therefore, do not survive the consistency check. Those pcis that pass the test do not entail  $\Delta p \wedge \Delta q$ . However, free choice permission can be derived if we assume a weaker notion of competence: if we take as context the set of states  $\mathcal{C}^+$  where besides  $[D]$ ,  $[4]$ , and  $[5]$  only  $[C_1]$  is valid but not  $[C_2]$  then  $\Box\Delta p$  passes the consistency check and entails  $\Delta p$  – and the same is true for  $\Box\Delta q$  and  $\Delta q$ .

As these considerations make clear, the ideas on which Gazdar’s work and the account introduced in section 2.3 are based are very similar. In the technical details, however, the approaches differ. For one thing, both proposals try to minimize the belief state of the speaker, however, they have different opinions about to which part of her beliefs this should be applied. The second discrepancy lays in the criteria the approaches apply to decide whether some belief state is a proper minimum. Below, both differences will be discussed in some detail.

Particularly the first difference is interesting for the discussion at hand. As we have seen in section 2.4.1, the approach introduced here takes too much of the belief state of the speaker to be relevant. Gazdar proposes a much more context-sensitive criterion to select relevant belief: relevant is what the speaker believes about the sentences that – in a very technical sense – the speaker is talking about: the subsentences of the uttered sentence. We can try and build this idea into the approach developed here. Maybe this way we can overcome the problem of overgeneration.

As already mentioned in a footnote in section 2.3.4.4 the order  $\preceq^+$  on which the notion of pragmatic entailment  $\models^+$  is based can be equivalently defined by comparing how many of a certain set of sentences the speaker believes.

**2.4.1.1. FACT.** Let  $\mathcal{L}^+ \subseteq \mathcal{L}$  be language defined by the BNF-form  $\chi_+ ::= p(p \in \mathcal{L}_{(0)}) \mid \chi_+ \wedge \chi_+ \mid \chi_+ \vee \chi_+ \mid \nabla p(p \in \mathcal{L}^0)$ . Then we have for  $s, s' \in \mathcal{C}$ :

$$s \preceq^+ s' \Leftrightarrow \forall \chi \in \mathcal{L}^+ : s \models \Box\chi \Rightarrow s' \models \Box\chi.$$

This representation of the order suggests a way how we can use Gazdar’s idea in our approach: instead of  $\mathcal{L}^+$  we take the sub-sentences of the uttered clause

---

<sup>33</sup>Gazdar himself never discussed this application of his formalization of Grice’s theory. In particular, it was not his intention to account for the free choice inferences this way.

as the set of sentences defining the order. Thus, let  $\mathcal{L}^+(\phi)$  be the set of subsentences of sentence  $\phi$ . We define:  $\forall s, s' \in \mathcal{S} : s \preceq^g s' \text{ iff}_{\text{def}} \forall \chi \in \mathcal{L}^+(\phi) : s \models \Box\chi \Rightarrow s' \models \Box\chi$ . This order can then be used to define a respective notion of entailment  $\models_S^{g+}$ . Applied to context  $\mathcal{C}$  this relation still accounts for the free choice inferences – when in the sentence interpreted *or* has wide scope over the modal expressions. Furthermore,  $\models_S^{g+}$  certainly predicts less false implicatures than does  $\models_S^+$ . For instance, for arbitrary and logical independent  $p, q, r \in \mathcal{L}^0$  we do not have  $p \vee q \models_S^{g+} \Diamond r \wedge \Diamond \neg r \wedge \Diamond \Delta r \wedge \Diamond \neg \Delta r$  (the same is true for  $\models_C^{g+}$ ). However, a restriction to subsentences does not completely solve the problem of overgeneration.  $\models_C^{g+}$  will predict wrongly for  $\Delta p \vee \Delta q$  the implicature  $\Diamond p$ .<sup>34</sup> Finally, there is also a conceptual problem with such an approach.  $\models_C^{g+}$  is still intended to describe a class of conversational implicatures and to formalize Grice's theory thereof. But what kind of Gricean motivation can be given for such restrictions of the inferences to subsentences of the sentence uttered?

To explain the second difference between Gazdar's approach and the one introduced in section 2.3 we should compare his approach with an even more Gazdarian variant of  $\models$ . As the reader may have noticed, he considers not only the sub-sentences of an uttered sentence to be relevant but also their negations. Let us define  $\mathcal{L}(\phi)$  as the closure of  $\mathcal{L}^+(\phi)$  under negation.  $\models_S^g$  is obtained by substituting the order  $\forall s, s' \in \mathcal{S} : s \preceq^g s' \text{ iff}_{\text{def}} \forall \chi \in \mathcal{L}(\phi) : s \models \Box\chi \Rightarrow s' \models \Box\chi$  in definition 2.3.1.

Intuitively, both Gazdar's description of clausal implicatures and  $\models^g$  do the same thing: making as many sentences  $\Diamond\chi$  true for  $\chi \in \mathcal{L}(\phi)$  as they can. However, the predictions made are different and this difference is due to the consistency check pcis have to pass before they become actual clausal implicatures. As we have said above, Gazdar predicts those pcis not to be generated that together with the context, the statement that the speaker knows  $\phi$  to hold, and some set of pcis satisfiable in the context lead to an inconsistency. What does  $\models_S^g$  predict in such a case? If  $\Diamond\chi$  for  $\chi \in \mathcal{L}(\phi)$  and  $\Diamond\Sigma = \{\Diamond\chi \mid \chi \in \Sigma\}$  for  $\Sigma \subseteq \mathcal{L}(\phi)$  are not jointly satisfiable in the set of states  $s \in S$  where  $\Box\phi$  is valid, while  $\Diamond\chi$  and  $\Diamond\Sigma$  separately are satisfiable in this context, then this means that there are states  $s_1 \models \Diamond\chi$  and  $s_2 \models \bigwedge \Diamond\Sigma$ , but that such states are incomparable with each other. For  $\phi$  to be honest there has to be a state  $s \in S$ ,  $s \models \Box\phi$  such that  $s \preceq^g s_1$  and  $s \preceq^g s_2$ . From this it follows that  $s \models \Diamond\chi \wedge \bigwedge \Diamond\Sigma$ . But this conjunction does not have any model. Thus  $\phi$  has to be dishonest. The pragmatic interpretation breaks down, no implicatures are generated. Gazdar's predictions are less severe. According to him, sets of sentences on which the knowledge of the speaker cannot be minimized without resulting in inconsistencies are not minimized. They are

---

<sup>34</sup>Though one (normally) infers from an utterance of *You may A or B* that the speaker takes the asserted deontic options also to be epistemically possible, this inference should rather be analyzed as part of the appropriateness conditions (presuppositions) of permissions (and obligations).

taken out, so to say, of the set of relevant sentences. The Gricean interpreter modeled by Gazdar is more tolerant with the speaker than the interpreter modeled here.

This has consequences for the cancellation properties for free choice inferences that both approaches predict. While both proposals model the same behavior of free choice inferences in case they conflict with the context or the semantic meaning of the utterance that triggers them, they differ in their predictions in case pcis are inconsistent with each other (given a particular context). Gazdar's approach cancels only those implicatures that give rise to the inconsistency. According to the account presented here in this case the speaker disobeys the Gricean Principle. Therefore, no implicatures are derived that would rely on taking the speaker to obey the principle. Empirical investigations have to show which of these positions makes the better predictions.

## 2.5 Conclusions

Why can we conclude on hearing (4) *You may go to the beach or go to the cinema* that the addressee may go to the beach and may go to the cinema? In this chapter we have proposed that this is due to pragmatic reasons. Free choice permission is explained as a conversational implicature that can be derived if the speaker is taken (i) to obey the Gricean maxim of quality and the first sub-clause of the maxim of quantity,<sup>35</sup> and (ii) to be competent on the deontic options, i.e. to know the valid obligations and permissions.

The proposal made in this chapter is not the first approach that tries to describe free choice permission as a conversational implicature.<sup>36</sup> What distinguishes it from others on the same line is that it provides a formally precise derivation of the free choice inferences. In particular, a formalization of the conversational implicatures that can be derived from the maxim of quality and the first sub-clause of the maxim of quantity is given. This part of the proposal essentially builds on work of Halpern & Moses (1984) on the concept of *only knowing*, generalized by van der Hoek et al. (1999, 2000).

A central feature of the presented account that distinguishes it from *semantic* approaches to the free choice inferences is that it maintains a simple and classical formalization of the semantics of English: modal expressions are interpreted as modal operators and *or* as inclusive disjunction. This has the advantage that the approach is free of typical problems that many semantic approaches to the free choice inferences have to face. For instance, when embedded under other logical operators, *or* behaves as if it means inclusive disjunction. Semantic approaches often cannot account for this observation (cf. Zimmermann 2000, Geurts 2005, Alonso-Ovalle 2004). Furthermore, because with such an approach to semantics

---

<sup>35</sup>These two maxims were combined in the Gricean Principle.

<sup>36</sup>See e.g. Kamp (1979), Merin (1992), van Rooij (2000).

$\Delta(p \vee q)$  and  $\Delta p \vee \Delta q$  are equivalent, the free choice inferences are predicted for both sentences, independent of whether *or* has wide or narrow scope with respect to the modal expressions. This allows us to account for the observation that free choice inferences can come with sentences like (13b) *You may take an apple or you may take a pear* as well. At the same time we are not forced to exclude a narrow scope analysis for *You may take an apple or a pear* (cf. Zimmermann 2000, Geurts 2005).

To summarize, we can conclude that the central goal of the work presented here, to come up with a formally precise pragmatic account to free choice permission, has been achieved. But there are still many questions concerning the behavior of free choice inferences that remain unanswered by the present approach.

The most urgent question is, of course, how to get rid of the countless unwanted pragmatic inferences the account predicts. Closer considerations in section 2.4.1 have suggested that this problem is a consequence of the fact that the approach incorporates only parts of Grice's theory of conversational implicatures. In particular, contextual relevance does not play any role. Future work has to reveal whether an extension of the approach in this direction helps to get rid of the problem of overgeneration.

An important topic that has received only marginal attention here was the question in how much the behavior of the free choice inferences forces us to adopt a pragmatic approach towards them. We have already noted that this is not easily answered. Much depends on the concept of pragmatic inferences that is adopted, on the classification of the data, and other theoretical decisions. In section 2.2 we have seen a series of arguments that speak in favor of a pragmatic approach. But the evidence is not as clear as this might suggest. Some observations argue rather for a semantic treatment of free choice inferences. For instance, the pragmatic inferences a sentence  $\phi$  comes with should be unaffected when in  $\phi$  semantically equivalent expressions (having roughly the same complexity) are exchanged. A pragmatic approach to the free choice inferences would thus predict, one may argue, that with *He may speak English or he may speak Spanish*, *He is permitted to speak English or he is permitted to speak Spanish* should also allow a free choice reading. This does not seem to be the case.<sup>37</sup> How serious a problem this is depends, of course, on the exact semantics assumed for *permit* and *may*. We cannot solve this issue here. The only point that we want to make is that the question whether the free choice inferences are semantic or pragmatic in character is essential for evaluating the pragmatic approach proposed here and, therefore, needs close attention in future work.

Another subject for future research is the additional and non-Gricean inter-

---

<sup>37</sup>This type of argument against a pragmatic account of the free choice inferences has been brought forward at different places in the literature. The particular example used here can be found, for instance, in Forbes (2003), as pointed out by one of the referees.

pretation principle – assuming the speaker to be competent – that is part of the approach. It is not the first time that such a principle is taken to be relevant for interpretation. In the literature of conversational implicatures there is even a long tradition in describing certain implicatures as involving such a competence assumption.<sup>38</sup> On the other hand, competence as formalized here is a very strong concept. One may wonder how reasonable it is to ascribe (by default) such a property to speakers. Therefore, it is important, for instance, to investigate whether the competence principle also shows itself in other areas of interpretation.

---

<sup>38</sup>One of the oldest references may be Soames (1982).



## Chapter 3

---

# Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation

(joint work with R. van Rooij)

### 3.1 Introduction

The central aim of this chapter is to find an adequate description of the particular way in which we often enrich the semantic meaning of answers.<sup>1</sup> To illustrate the phenomenon, consider the following dialogue.

- (16) Ann: Who passed the examination?  
Bob: John and Mary.

In many contexts Bob's answer is interpreted as *exhausting* the predicate in question, hence, as stating not only that John and Mary passed the examination, but also that these are the only people that did. This reading is called the *exhaustive interpretation* of answers (see e.g. Groenendijk & Stokhof (1984), von Stechow & Zimmermann (1985)) which we will study in this chapter.<sup>2</sup>

The term *exhaustive interpretation* has not only been used in connection with the interpretation of answers. Aspects of the meaning<sup>3</sup> of sentences containing

---

<sup>1</sup>This chapter has been published as 'Pragmatic meaning and non-monotonic reasoning. The case of exhaustive interpretation' 2006 in *Linguistics and Philosophy*, **29**(2): 205-250. The article is reprinted here with the kind permission from Springer Science and Business Media.

<sup>2</sup>As will become clearer in section 3.2, we will treat the particular reading exemplified in (16) as only a special case of exhaustive interpretation.

<sup>3</sup>In this chapter *meaning* will be used as referring to all the information conveyed by an utterance in a particular context.



*only* (compare *Only John and Mary passed the examination*), cleft constructions (*It was John and Mary who passed the examination*) and free intonational focus (*[John and Mary]<sub>F</sub> passed the examination*), for instance, have been characterized in this way as well. In this chapter, however, we will limit ourselves to a description of the exhaustive interpretation of answers. We will discuss semantic analyses of these other constructions only insofar as they have to do with problems that arise with the exhaustive interpretation of answers as well.

In their dissertation from 1984, Groenendijk & Stokhof proposed a very promising approach to the exhaustive interpretation of answers. We will introduce this approach in section 3.3 and discuss its merits. However, Groenendijk & Stokhof's (1984) description of exhaustive interpretation also faces certain shortcomings. The main goal of the remaining sections is to provide the necessary changes and adaptations to overcome these limitations.

In section 3.4 we will discuss the close relation between Groenendijk & Stokhof's (1984) approach, McCarthy's (1980, 1986) theory of predicate circumscription, and the latter's model-theoretic variant: interpretation in minimal models. We will then switch to a description of exhaustive interpretation as interpretation in minimal models and show that this already allows us to address some of the problems Groenendijk & Stokhof (1984) have to face.

In section 3.5 another modification is added: we will combine the new approach with dynamic semantics. Given the developments in semantics during the last 20 years, this is an alteration of the original static approach of Groenendijk & Stokhof (1984) that would have been necessary anyway. It will turn out that it solves some problems, already discussed by Groenendijk & Stokhof (1984) themselves, concerning, for instance, the interaction of exhaustive interpretation and the semantics of determiners.

In section 3.6 we will address the context-dependence of exhaustive interpretation. As will be illustrated in section 3.2, exhaustive interpretation can come in other forms than the reading we discussed for example (16). We will argue that this should be explained by taking a contextual parameter of relevance into account.

In the final section we will go beyond our primary aim to provide an adequate description of exhaustive interpretation. The need of such a description arises because standard semantics cannot handle the phenomenon. That means that if we want to maintain standard semantics exhaustive interpretation cannot be explained as a semantic phenomenon. But where does it come from if not from semantics? One answer to this question that seems to be particularly attractive is to analyze it as a Gricean conversational implicatures. We will sketch a formalization of parts of Grice's theory brought forward by Schulz (2005) and van Rooij & Schulz (2004). It can be shown that when combined with a principle of competence maximization, this formalization indeed accounts for exhaustive interpretation (as described in section 3.4).

## 3.2 The phenomenon

Before we can start thinking about how to formulate a general and precise description of the exhaustive interpretation of answers, we first need to get a clearer picture of what we actually have to describe. Therefore, this section is devoted to a closer investigation of the properties of exhaustive interpretation.

### 3.2.1 Interaction with the semantic meaning of the answer

The first thing to notice is that an exhaustive interpretation does not always completely resolve the question the answer addresses. Consider, for instance, example (17) (all sentences discussed in this section should be understood as answers to the question *Who passed the examination?*).

(17) Some female students.

This answer can be interpreted exhaustively as stating that just a few students passed the examination and that they are all female. However, also on this reading the answer does not identify the students that passed the examination and therefore does not resolve the question.<sup>4</sup> Hence, even though the exhaustive interpretation strengthens the standard semantic meaning of the answer – and therefore makes them arguably better answers – it does not turn all answers into resolving ones. This point is also nicely illustrated with the following example.

(18) John or Mary.

On its exhaustive interpretation this answer states that either only John or only Mary exhaust the set of people who passed the examination. Again, in most contexts this information will not fully resolve the question asked.<sup>5</sup> Another point that should be noticed in connection with this example is that its exhaustive interpretation is not completely described by taking the exclusive interpretation of *or*.<sup>6</sup> One would miss the additional inference of the exhaustive reading that no-one else besides Mary and John passed the examination.

---

<sup>4</sup>This is true, in particular, for the notion of resolving questions introduced by Groenendijk & Stokhof (1984). According to them, a question is resolved if the extension of the question-predicate is fully specified. One may argue that resolvedness should also depend on the information the questioner is interested in when asking her question (see Ginzburg (1995) and van Rooij (2003) for proposals along these lines). However, even if one modifies the notion of resolving questions accordingly, it will not be the case that the exhaustive interpretation of an answer like (17) will always resolve the question.

<sup>5</sup>Sometimes, however, a disjunction can be resolving. Consider, for instance, (i) reading Ann's utterance as a polar question.

(i) Ann: Did Mary or John pass the examination?  
 Bob: Yes, Mary or John passed the examination.

<sup>6</sup>Under the exclusive interpretation of *or*, *A or B* is true iff one of the disjuncts is true but not both.

Careful attention should be paid also to the way exhaustive interpretation interacts with the semantics of determiners. Compare, for instance, (19) and (20).

(19) Three students.

(20) At least three students.

The exhaustive interpretation of (19) allows us to conclude that not more than three students passed the examination. (20), however, cannot be read in this way.<sup>7</sup> So there is a difference between (19) and (20) that exhaustive interpretation is sensitive to. However, it will not be adequate to propose that *at least* simply cancels an exhaustive interpretation. (20) can give rise to the inference that nobody besides students passed the examination, and, thus, can show effects of exhaustification.

For (21), just as for (20), we will not infer a limitation on the number of students that passed the examination, if it is interpreted exhaustively.

(21) Students.

Notice that nevertheless it *can* be concluded that, besides students, no one else passed the examination. Thus, also in this case certain effects of exhaustive interpretation are present. In contrast to (21), the exhaustive interpretation of (22) implies additionally that not all students passed the examination. So, again, something distinguishes (21) and (22) with respect to exhaustive interpretation.

(22) Most students.

How can these observations be explained? We will propose in section 3.5 that exhaustivity is sensitive to the different dynamic semantics of the answers and this leads to the different interpretations.<sup>8</sup>

### 3.2.2 The context-dependence of exhaustivity

The examples discussed above show how exhaustive inferences change depending on the answer given. Interestingly enough, even the same answer (following the same question form) can give rise to different exhaustive interpretations in different contexts. First of all, it seems that sometimes answers should not be interpreted exhaustively at all. A typical example is the dialogue given in (23).

---

<sup>7</sup>We only discuss here *at least* as a modifier of the numeral. The occurrence of *at least* in (20) can also be read as particle, with a syntactical behavior similar to *even*. This use is not discussed in the present chapter. Readers who have problems getting the exhaustive interpretation for (20) should try *John and at least three of his friends passed the examination*.

<sup>8</sup>Some of the inferences attributed in this section to exhaustive interpretation are standardly analyzed as conversational implicatures. This is no accident, as we see it. In section 3.7 we will discuss the relation between exhaustive interpretation and conversational implicatures in some detail.

- (23) Ann: Who has a light?  
 Bob: John.

Here, Bob's answer is normally not understood as *John is the only one who has a light*. Instead, it seems that no information other than its semantic meaning is conveyed. We call this interpretation of answers the *mention-some* reading, while we will refer to the one discussed until now as exhaustive interpretation as the *mention-all* reading. It appears that mention-some readings occur precisely in those contexts where the questioner is intuitively not interested in the exact specification of the question predicate and the semantic meaning of the answer already provides her with all the information she needs.<sup>9</sup>

Aside from the *mention-all* and *mention-some* readings, there also seem to be situations with *intermediate* exhaustive interpretations. In these cases some of the typical inferences of mention-all readings are allowed, but not all of them.

Perhaps the best known limitation is *domain restriction*. There are contexts in which an answer to a question with question-predicate *P* specifies those and only those individuals that have property *P* - but only for a *subset* of all objects to which *P* may apply. Imagine Mr. Smith asking one of his employees:

- (24) Mr. Smith: Who was at the meeting yesterday?  
 Employee: John and Mary.

There is a reading of this answer implying that John and Mary are the only *employees* of Mr. Smith who were at the meeting yesterday. There may have been others besides employees of Mr. Smith, but nothing is inferred about them. For the choice of interpretation it seems to be relevant, again, what is commonly known about the information Mr. Smith is interested in. Suppose, for instance, that it is mutually known that Mr. Smith would like to know whether one of his rivals from other companies was at this meeting. Then one would infer from (24) that John and Mary are the only *rivals* of Mr. Smith who were at the meeting yesterday.

Exhaustive interpretation is limited in other ways in so-called *scalar readings* of answers (cf. Hirschberg, 1985). As in the example above, also here exhaustivity seems to apply only to parts of the question-predicate. Imagine Ann and Bob playing poker.

---

<sup>9</sup>It is important to distinguish mention-some readings from a different interpretation an answer can get, for instance, if the speaker adds, for instance, *as far as I know*. In contrast to mention-some readings, in the latter cases the information one receives is not exhausted by the semantic meaning of the answer. Instead it is additionally inferred that the information given in the answer exhausts the *knowledge* of the speaker. Of course, this latter reading can also occur if the questioner is interested in a full specification of the question-predicate. See section 3.7 for more discussion.

(25) Ann: What cards did you have?

Bob: Two aces.

Here, Ann will interpret Bob's answer as saying that he did not have three aces or two additional kings (a double pair wins over a single one). Still, the answer intuitively leaves open the possibility that Bob additionally had, for example, a seven, a nine, and the king of spades. Just as in the previous case, Ann's interest in information here is different from the case in which an answer gets a mention-all reading. She is not interested in the exact cards that Bob had. She wants to know, however, how *good* (with respect to an ordering relation induced by the poker rules) Bob's cards were. And the scalar reading tells her that Bob did not have additional cards that would raise this value.

To give a final example of a context where the force of exhaustive interpretation seems to change depending on the context, consider (26).

(26) Ann: How far can you jump?

Bob: Five meters.

If it is commonly known that Ann wants to have precise information about Bob's jumping capacities, the exhaustive interpretation of his answer will imply that he cannot jump a centimeter further than 5 meters. If, however, a rough indication is sufficient, one may infer just that he cannot jump 6 meters. This illustrates how – depending on the needs of the questioner – exhaustive interpretation can select the domain of the question-predicate with different degrees of fine-grainedness.

In this subsection we have discussed some examples where an answer does not obtain the strong interpretation that is traditionally associated with the name *exhaustive interpretation*. Sometimes only parts of the mention-all reading were observed, sometimes nothing was added to the semantic meaning of the answer at all. But in all cases the contextual parameter on which the strength of the exhaustive interpretation depends seems to be what is commonly known to be relevant for the questioner. If the questioner is known not to be interested in certain information, then it will not be provided by the exhaustive interpretation of the answer. For instance, in a typical context where (23) is used it is clear that for the questioner, it is sufficient to know of somebody who has a light that she has a light. This interest is fully satisfied with the semantic meaning of the answer given by Bob. We will take this observation seriously and describe in section 3.6 exhaustive interpretation as depending on the information the questioner is interested in. It will turn out that in this way we can account for domain restrictions, granularity effects, scalar readings, as well as the mention-some interpretation.

### 3.2.3 Other types of questions

The examples discussed so far all were answers to some *wh*-question where the question-predicate is of type  $\langle s, \langle e, t \rangle \rangle$ . The exhaustive interpretation is, however,

not restricted to this class of answers. There are questions of other types whose answers also seem to show exhaustiveness effects. For instance, there is a well-known tendency to interpret conditional answers to polar questions, exemplified in (27), as bi-conditional.

(27) Ann: Will Mary win?

Bob: Yes, if John doesn't realize that she is bluffing.

Thus, one infers that Mary will win just in case John does not realize that she is bluffing. Intuitively, in this reading the same thing is going on as in the cases of exhaustive interpretation discussed so far: the worlds where Mary will win are taken to be exhausted by those where the antecedent of Bob's answer is true. Therefore, it seems reasonable to expect that a convincing approach to exhaustive interpretation should be able to deal with this observation as well.

---

Various approaches to exhaustive interpretation already exist in the literature. However, the domain of application differs markedly from theory to theory. As far as we know, none of the existing approaches can account for all the observations discussed above. Moreover, none of these theories gives a satisfactory explanation for why the scope of exhaustive interpretation should be restricted to those cases that they can actually handle.

In this chapter a unified approach to the exhaustive interpretation of answers is presented which is able to deal with the whole list of examples discussed so far. This account essentially builds on a description of exhaustive interpretation proposed by Groenendijk & Stokhof (1984) (abbreviated by G&S). We will therefore start by discussing their work.

There is one final point that should be made clear. The reader may have noticed that some of the inferences attributed here to exhaustive interpretation are standardly analyzed as conversational implicatures. As we will try to argue in section 3.7, this is to be expected because exhaustive interpretation is by itself a conversational implicature. But then one may wonder what the relevance of this chapter is, for we already have Grice's theory to account for conversational implicatures. However, the well-known problem of this theory is that it does not make clear predictions, and although many people have tried we are not aware of any fully satisfying formalization of Grice's proposal. Therefore, if we want to account in terms of it for exhaustive interpretation, we need to provide at least a partial formalization of Grice's theory. This will be the topic of section 3.7. But to evaluate whether this formalization indeed accounts for exhaustive interpretation, and also as starting point for theories that do not agree with our opinion that exhaustive interpretation is a conversational implicature, one first needs an adequate description of this rule of interpretation. The sections 3.4 to 3.6 of this chapter will deal with this issue.

But let us start with discussion the classical approach to exhaustive interpretation of Groenendijk & Stokhof (1984).

### 3.3 Groenendijk and Stokhof's proposal

Assume that  $W$  is a class of models (possible worlds) for our language and let  $[\phi]$  denote the intensional semantic meaning of expression  $\phi$ . Hence,  $[\phi]$  is a function mapping elements  $w$  of  $W$  on the extension of  $\phi$  in  $w$  (in case  $\phi$  is a sentence we use  $[\phi]^W$  to denote the set of models in  $W$  where  $\phi$  is true). Groenendijk & Stokhof (1984) propose to describe the exhaustive interpretation of answers to *Who*-questions by the operation  $exh_{GS}$  taking as arguments the generalized quantifier denoted by the term answer  $T$  and the property denoted by the question-predicate  $P$ .<sup>10</sup>

**3.3.1. DEFINITION.** (The exhaustivity operator of Groenendijk & Stokhof)

$$exh_{GS}([T], [P]) =_{def} \lambda w. [T]([P])(w) \wedge \neg \exists \mathcal{P}' : [T](\mathcal{P}')(w) \wedge \mathcal{P}'(w) \subset [P](w)$$

Set-theoretically, the above formula applied to a generalized quantifier and a property allows the property only to select the *minimal* elements of the generalized quantifier. To illustrate, assume that Bob's response to Ann's question *Who passed the examination?* is *John*. Analyzed as a general quantifier *John* denotes  $\lambda w \lambda \mathcal{P}. \mathcal{P}(w)(j)$ , which is true of some property  $\mathcal{P}$  if in every world  $\mathcal{P}$  denotes a set containing  $j$ . Applying  $exh_{GS}$  to this function turns it into a generalized quantifier that is true of  $\mathcal{P}$  if in every world it denotes the *minimal* set containing  $j$ , which is the set  $\{j\}$ . Thus, it is correctly predicted that by exhaustive interpretation we can conclude from the answer *John* that  $\{j\}$  is the set of individuals that passed the examination.

Reading term answers as generalized quantifiers in combination with the exhaustivity operation defined above allows us to account for the interpretation effects observed in examples as (16), (17), (18), (19) and (22) discussed in sections 3.1 and 3.2. Actually, G&S can do even more. They show that the above stated operator for terms can be generalized easily to  $n$ -ary question predicates.<sup>11</sup>

Although these results are very appealing, G&S's exhaustivity operator has still been criticized. For instance, Bonomi & Casalegno (1993) have argued that G&S's

<sup>10</sup>As mentioned in G&S (1984), this operator has much in common with Szabolcsi's (1981) interpretation rule for *only*. Though similar in content, G&S's version provides the more transparent formulation.

<sup>11</sup>Their general exhaustivity operator for  $n$ -ary terms looks as follows:

$$exh_{GS}^n([T_n], [P_n]) = \lambda w. [T_n]([P_n])(w) \wedge \neg \exists \mathcal{P}'_n : [T_n](\mathcal{P}'_n)(w) \wedge \mathcal{P}'_n(w) \subset [P_n](w).$$

analysis is rather limited because it can be applied only to noun phrases. To account for examples in which *only*<sup>12</sup> associates with expressions of another category, they argue that we have to make use of events. We acknowledge that the use of (something like) events might, in the end, be forced upon us. But perhaps not exactly for the reason they suggest. Crucial for G&S's analysis is that (ignoring the intensional parameter) their exhaustivity operator is applied to objects of type  $\langle\langle\phi, t\rangle, t\rangle$ . It is normally assumed that noun phrases denote generalized quantifiers of type  $\langle\langle e, t\rangle, t\rangle$ , which means that denotations of noun phrases are in the range of the exhaustivity operator. However, it is also standardly assumed that an expression of *any* type  $\phi$  can be *lifted* to an expression of type  $\langle\langle\phi, t\rangle, t\rangle$  without a change of meaning. But this means that – after type-lifting – G&S's exhaustivity operator can be applied to the denotation of expressions of any type, and there is no special need for events.<sup>13</sup>

Although Bonomi & Casalegno's (1993) criticism does not seem to apply, G&S's analysis faces some other limitations. First, it is quite obvious (and has been noticed by themselves) that they cannot account for mention-some readings, domain restriction, granularity effects, and scalar readings (see section 3.2.2). This is inevitable given the functionality of  $exh_{GS}$ . By taking only the semantic meaning of the predicate of the question and the term in the answer as arguments,  $exh_{GS}$  is too rigid to account for differences that can occur involving the same question-predicate and the same answer. The limited functionality of the operation  $exh_{GS}$  also seems to be responsible for other problems of the approach (called *the functionality problem* by Bonomi & Casalegno (1993)). Because G&S assign to (19) *Three students* and (20) *At least three students* the same meaning,  $exh_{GS}$  predicts for these pairs of answers the same exhaustive interpretation. However, as discussed in section 3.2.1, intuitively the interpretations differ.<sup>14</sup> Something similar is the case for answers like (18) *John or Mary* and (28).

(28) John or Mary or both.

Standard semantics takes both answers to be equivalent, but their exhaustive interpretation differs. While (18) implies that John and Mary did not pass the examination, this is not true for (28).  $Exh_{GS}$  predicts the exclusive reading in both cases.

---

<sup>12</sup>They discuss  $exh_{GS}$  as a description of the semantic meaning of *only*, but their criticism applies with the same force to  $exh_{GS}$  as a description of the exhaustive interpretation of answers.

<sup>13</sup>The reason that we still might need (something like) events is that for questions as *What did you do last summer?* a possible-world approach may not provide enough fine-structure to properly describe the meaning of the question-predicate, thus, the set of 'things' one did last summer. Making use of events may be one way to achieve this required fine-grainedness. But this is not a problem of G&S's approach to exhaustive interpretation but rather for the general conception of meaning in which this proposal is situated.

<sup>14</sup>In both cases the application of  $exh_{GS}$  implies that not more than three students passed the examination. This problem has also been noted by G&S themselves.



The next problem (discussed in G&S, pp. 416-417) concerns the way  $exh_{GS}$  operates on its arguments. If we allow for group objects, interpret *[passed the examination]* as a distributive predicate<sup>15</sup> and *[John and Mary]* in (29) as the quantifier  $\lambda w.\lambda \mathcal{P}.\mathcal{P}(w)(j \oplus m)$ , then G&S predict that on the exhaustive interpretation of (29) *[passed the examination]* denotes the set  $\{j \oplus m\}$ .

- (29) Ann: Who passed the examination?  
 Bob: John and Mary.

Because *[passed the examination]* is distributive, this cannot be fulfilled in any world: there can be no model  $w$  of the language where  $j \oplus m \in [passed\ the\ examination](w)$  but  $j \notin [passed\ the\ examination](w)$ . Hence, when applying  $exh_{GS}$  to (29) Bob's answer is interpreted as the absurd proposition. This is inadequate given that (29) can be interpreted straightforwardly in an exhaustive way.<sup>16</sup>

Finally, negation is a problem for  $exh_{GS}$ . Apply, for instance, this operation to Bob's answer in (30).

- (30) Ann: Who passed the examination?  
 Bob: Not John.

Then Bob's response is interpreted as implying that *nobody* passed the examination: the smallest extension of predicate *passed the examination* such that the answer is true is the empty set. This is clearly not a possible reading for this answer.

The aim of this chapter is to overcome the problems discussed above. We claim that this can be done without radically changing the basic idea behind G&S's exhaustivity operator.

What do we understand this basic idea to be? According to G&S, to interpret an answer exhaustively means to minimize the question-predicate of the answer: from the fact that the answerer did not claim that a certain object has property  $\mathcal{P}$  it is inferred that the object does not have property  $\mathcal{P}$ . Thus, the hearer makes the absence of information meaningful. She interprets it as negation. This we take to be an essentially correct perspective on what exhaustive interpretation is about.

---

<sup>15</sup>A predicate  $P$  with domain  $D$  is distributive in a set of models  $W$  if for all  $w \in W$ ,  $(\forall x, y \in D)([P](w)(x) \wedge [P](w)(y) \leftrightarrow [P](w)(x \oplus y))$ .

<sup>16</sup>There is a solution to this problem, already sketched by G&S, *ibid.* For independent reasons one is driven to allow the interpreter to choose freely between a distributive and non-distributive reading for predication to plural objects. If one additionally assumes that distributive predicates allow only for the second reading, (29) is interpreted as  $\forall x \leq j \oplus m : P(x)$ . Minimization of  $P$  relative to this answer does not give rise to complications. Later on (section 3.4.2) we will propose another solution. It has the advantage to carry over to a different kind of problem that the proposal sketched here cannot capture.

However, G&S were not aware of the fact that this reasoning pattern – negation as failure – was starting to get a lot of attention in artificial intelligence as well. It lead (among other things) to the development of a whole new branch of logic: *non-monotonic* logic. When we now try to improve on the proposal of G&S we can build on the work done in this area.<sup>17</sup>

## 3.4 Exhaustivity as Predicate Circumscription

### 3.4.1 Predicate Circumscription

Only a few years before Groenendijk and Stokhof came up with their description of exhaustive interpretation, McCarthy impressed the artificial intelligence community by introducing *Predicate Circumscription*, one of the first formalisms of non-monotonic logic. McCarthy’s goal was to formalize common sense reasoning. More specifically, Predicate Circumscription was intended to solve the *qualification problem*: if we would use classical logic to derive every-day conclusions, we would need an “impracticable and implausible” (McCarthy, 1980, p. 145) number of qualifications in the premises. For instance, if one wants to predict that if we would throw our computers out of our windows, they would smash on Nieuwe Doelenstraat, one would have to specify that gravitation will not stop working, the computers will not develop wings and fly away etc. - in short: nothing extraordinary will happen. The solution McCarthy proposes is to strengthen the inferences one can draw from a theory by adding to the premises a *normality* assumption. It says that nothing abnormal is the case that is not explicitly mentioned in the theory. Or, to restate it somewhat more abstractly, the extension of certain predicates (the abnormality predicates) is restricted to those and only those objects that are explicitly stated by the premises to be in the extension. To come back to the example above, if there is no explicit information about abnormalities in the gravitation of the earth the normality assumption adds the premise that the gravitation is working as normal. Thus, abnormality is negated as failure.

McCarthy (1986) formalizes this idea<sup>18</sup> by defining a syntactic operation on a sentence (the premise) that maps it to a new second-order sentence (the premise plus the normality assumption) in the following way.

---

<sup>17</sup>A question one often hears in this context is *Do we really need non-monotonic logic?*. Indeed, we do. Non-monotonicity is simply a property of exhaustive interpretation. Therefore, no matter how one describes exhaustive interpretation, it will also be a property of the description. Recall that reasoning is non-monotone if certain inferences might be given up under the presence of more information. It is easy to see that this holds for exhaustive interpretation. From the answer *John* we can conclude that Mary did not pass the examination. This inference is lost when the speaker also tells us that Mary passed as in (29).

<sup>18</sup>This is a simplified version of his formalization.

**3.4.1. DEFINITION.** (Predicate Circumscription)

Let  $A$  be a second-order formula and  $P$  a predicate of some language  $\mathcal{L}$  of predicate logic. Then the circumscription of  $P$  relative to  $A$  is the formula  $\text{CIRC}(A, P)$  defined as:

$$\text{CIRC}(A, P) := A \wedge \neg \exists P' : A[P'/P] \wedge P' \subset P,$$

where  $A[P'/P]$  describes the substitution of all free occurrences of  $P$  in  $A$  by  $P'$ .

Looking at this formalization of Predicate Circumscription, our reader will immediately recognize the following striking fact: G&S's exhaustivity operation is – roughly speaking – just an instantiation of McCarthy's predicate circumscription! The circumscribed predicate is now the question-predicate, and the circumscription is relative to the sentence one gets by combining term-answer and question-predicate - or simply the sentential answer. This important parallelism was first noticed, as far as we know, by Johan van Benthem (1989).<sup>19</sup>

Predicate circumscription has a model-theoretic pendant: interpretation in *minimal models*. First, the model-theory for classical logic is enriched by defining an order on the set of models  $W$ : a model  $v$  is said to be more minimal than a model  $w$  with respect to some predicate  $P$ ,  $v <_P w$ , in case they agree on everything except the interpretation they assign to  $P$  and it holds that  $[P](v) \subset [P](w)$ . It can be shown that if  $W$  is the class of all models the  $P$ -minimal models of a theory  $A$ , hence the set  $\{w \in [A]^W \mid \neg \exists v \in [A]^W : v <_P w\}$ , are exactly the models where the circumscription formula  $\text{CIRC}(A, P)$  holds.<sup>20</sup>

---

<sup>19</sup>There are certain differences between  $exh_{GS}$  and  $\text{CIRC}$  that should be mentioned. First, G&S took  $exh_{GS}$  to be a description of an operation on semantic representations while  $\text{CIRC}(A, P)$  is an expression in the object language. Second,  $\text{CIRC}$  takes as arguments a predicate and a sentence, while  $exh_{GS}$  applies to the predicate and the sentence without the predicate.  $\text{CIRC}$ , therefore, relies on less syntactic information. But, as Ede Zimmermann pointed out to us, it looks as if there are cases where exhaustive interpretation relies on this information. Consider, for instance, the answer *Men that wear a hat* to a question *Who wears a hat?*, where the question-predicate  $P$  appears in the term answer part  $T$ . The circumscription of  $A = T(P)$  w.r.t.  $P$  minimizes  $P$  in all occurrences of  $A$  and interprets the answer as implying that nobody wears a hat – which is obviously wrong.  $Exh_{GS}$  only minimizes occurrences of  $P$  outside  $T$  and correctly predicts that exactly those people wear a hat that are men that wear a hat. In this chapter we will assume that the question-predicate does not occur in the term answer part.

<sup>20</sup>This set of minimal models can be described relative to a set of alternatives of  $A$ ,  $\text{Alt}(A)$ , as well. If we say that  $v <_{\text{Alt}(A)} w$  if and only if  $v$  is exactly like  $w$  except that  $\{B \in \text{Alt}(A) \mid v \in [B]^W\} \subset \{B \in \text{Alt}(A) \mid w \in [B]^W\}$ , we can define the following set of minimal models:  $\{w \in [A]^W \mid \neg \exists v \in [A]^W : v <_{\text{Alt}(A)} w\}$ . This set is the same as the one described in the main text if we define  $\text{Alt}(A)$  as follows:  $\{P(a) \mid d \in D \ \& \ a \text{ is the name of } d\}$ , and assume that every individual has a unique name. A similar notion of alternatives is used in alternative-semantics approaches to the meaning of *only* (e.g. Rooth 1996). For more discussion see van Rooij & Schulz (2007).

This formulation of predicate circumscription – as interpretation in minimal models – is not a stranger to linguists. The Lewis/Stalnaker approach to counterfactuals (see, for instance, Lewis (1973)) also makes use of it. The application at hand differs mainly in the way the *order* is defined.

### 3.4.2 The basic setting

It is this later, model-theoretic formulation that we will use to describe exhaustive interpretation. Here comes our basic definition.

#### 3.4.2. DEFINITION. (Exhaustive interpretation - the basic case)

Let  $A$  be an answer given to a question with question-predicate  $P$  in context  $W$ . We define the exhaustive interpretation  $exh_{std}^W(A, P)$  of  $A$  with respect to  $P$  and  $W$  as follows:

$$exh_{std}^W(A, P) \equiv \{w \in [A]^W \mid \neg \exists v \in [A]^W : v <_P w\}$$

To illustrate the working of this interpretation function, let us go back to example (27) here repeated as (31).

(31) Ann: Will Mary win?

Bob: Yes, if John doesn't realize that she is bluffing.

In this case the question-predicate  $P = \text{Mary will win}$  is of arity 0.<sup>21</sup> But this means that  $v <_P w$  iff  $v$  is exactly like  $w$ , except that whereas  $w$  makes  $P$  true,  $v$  makes it false. Now it can be checked that  $exh_{std}^W(A \rightarrow P, P)$  is true only in those worlds where either both  $A$  and  $P$  are true, or both  $A$  and  $P$  are false. Worlds where  $A$  is false and  $P$  true are ruled out because there are other worlds that verify  $A \rightarrow P$ , but do not make  $P$  true (worlds where both  $A$  and  $P$  are false) and, hence, are more minimal. The possibility that  $A$  is true and  $P$  is false is excluded by the semantic meaning of the answer. Thus, by applying  $exh_{std}$  the conditional answer gets the desired bi-conditional reading.<sup>22</sup>

The change from G&S's approach to the one given in definition 3.4.2 is rather subtle – mainly one of perspective. But, as we will see in the rest of the chapter, this model-theoretic description of exhaustive interpretation proves to easily admit the amendments we have to make to deal with the limitations of G&S's approach. It also allows us to improve on  $exh_{GS}$  directly. Remember our earlier discussion of applying the rule of exhaustive interpretation to distributive predicates (enriching the domain with group objects). We discussed it using our very first example, repeated here as (32).

<sup>21</sup>We assume that the extension of an  $n$ -ary predicate  $P^n$  in world  $w$  is the set of  $n$ -ary tuples that verifies sentence  $P^n(\vec{x})$  in  $w$ . If  $P^0$  is true in  $w$ , it denotes  $\{\langle \rangle\}$ , otherwise  $\emptyset$ .

<sup>22</sup>This prediction is, of course, already made by G&S's operation  $exh_{GS}$ .

- (32) Ann: Who passed the examination?  
 Bob: John and Mary.

Let us calculate once more the exhaustive interpretation of Bob's answer, but now using  $exh_{std}$ . Again, Bob is taken to be talking about a plural object  $j \oplus m$ . To determine  $exh_{std}^W(P(j \oplus m), P)$  we first eliminate all worlds where Bob's answer is false. Then, we select those worlds where the extension of the question-predicate  $P$  is minimal. At first one may think that these are the worlds where the extension of *passed the examination* contains only  $j \oplus m$ . However, such worlds do not exist. The predicate is distributive and already G&S account for this by letting meaning postulates impose restrictions on the class of proper models. But then, the smallest extensions  $P$  can receive in worlds where Bob's answer is true are such that besides the plural object  $j \oplus m$  also  $j$  and  $m$  are in the extension of question-predicate *passed the examination*. Thus, we obtain the right result.

But why can we solve this problem just by taking  $exh_{std}$  instead of  $exh_{GS}$ ? Did we not claim above that  $exh_{GS}([T], [P])$  is roughly the same as  $CIRC(T(P), P)$  and the latter (more particularly  $[CIRC(T(P), P)]^W$ ) is equivalent to  $exh_{std}^W(T(P), P)$ ? Well, one has to be careful. Remember that the latter equivalence only holds if  $W$  is the class of *all* models. Meaning postulates impose restrictions on  $W$ .  $Exh_{std}$  is sensitive to these restrictions because they influence the set possible worlds it quantifies over.  $Exh_{GS}$ , however, quantifies locally over alternative extensions for the question-predicate. It does not check whether these alternatives are realized in some world. Only if the meaning postulates are taken to be part of the answer,  $exh_{GS}$  and  $CIRC$  are forced to respect them and predict correctly.

To sum up, distributive predicates show that circumscribing just the answer may not be enough. The exhaustive interpretation is sensitive to information available in the context set  $W$ , in particular to meaning postulates. Because  $exh_{std}$  quantifies over  $W$  it can immediately account for this dependence.<sup>23</sup>

Actually, there are even more striking examples in favor of a notion of exhaustivity which respects meaning postulates and they do not rely on particular premises such as the group analysis of Bob's answer in (32). For instance, the proposed formalization also allows us to account for some puzzles connected with the meaning of *only*. For limitations of space, however, we cannot discuss this issue in detail here.<sup>24</sup>

---

<sup>23</sup>The variable  $W$  makes our interpretation function very context dependent. If  $W$  is understood as the respective common ground then all the information presented there will influence what counts as a minimal model in a particular context. It still has to be seen to what extent exhaustive interpretation is context sensitive in this sense. See also the discussion in section 3.7.

<sup>24</sup>One particularly famous example that this approach can account for is the following from Kratzer (1989).

- (i) Bob: I only [painted a still-life]<sub>F</sub>.
- (ii) Lunatic: No. You also [painted apples]<sub>F</sub>.

For a closer discussion see van Rooij & Schulz (2007).

### 3.5 Exhaustivity and dynamic semantics

Another problem of G&S's approach that we discussed in section 3.2.1 is that it makes incorrect predictions for answers like (20) and (21), here repeated as (33b) and (33c).

- (33) (a) Three Students.  
 (b) At least three students.  
 (c) Students.  
 (d) Most students.

As we pointed out earlier, it is standardly assumed in generalized quantifier theory (adopted by G&S) that *three students* has the same semantic meaning as *at least three students*. Because the operation  $exh_{GS}$  (the same is true for  $exh_{std}$ ) takes only the semantic meaning of the answer and the question-predicate into account, it predicts for (33a) and (33b) the same readings. However, the exhaustive interpretation of the first answer gives rise, intuitively, to an *at most* inference, while the exhaustive interpretation of the latter does not. Something similar has been observed comparing (33c) and (33d). Thus, there is a difference between these answers exhaustive interpretation is sensitive to which  $exh_{GS}$  (and  $exh_{std}$  as well) fails to observe.

Different perspectives are possible on this dilemma. An interesting proposal is made by Zeevat (1994), who incorporates the *at most* inference (33d) comes with in the semantics of *most*. In this chapter, however, we stick to the traditional analysis of this determiner.<sup>25</sup> Others have proposed that expressions containing *at least* or bare nominals should not be interpreted exhaustively. However, as observed in section 3.2.1, also for these expressions we observe *some* exhaustivity effect. Hence, total absence of exhaustification is no option. We will propose instead that exhaustive interpretation *does* take place but that it will not give rise to the *at most* inference.

There is a difference between, for instance, (33c) and (33d) that can be made responsible their unequal exhaustive meanings. But – or so we propose – it is a difference in their semantic meaning.<sup>26</sup> The answers diverge in their dynamic discourse contribution. In consequence, to be able to make the correct predictions we have to adopt a dynamic perspective on semantics and describe exhaustive interpretation as an operation that is sensitive to dynamic information.

We will not introduce full-blooded dynamic semantics but restrict ourselves to some of its essential features, leaving the exact implementation to the reader's favorite dynamic theory. We assume a dynamic interpretation function that maps an information state  $\sigma$  and a sentence  $\phi$  to the new information state  $\sigma[\phi]$ . An

<sup>25</sup>Still, our final explanation will have some similarity with Zeevat's proposal.

<sup>26</sup>Thus, in this case it is not the functionality of  $exh_{GS}$  (or  $exh_{std}$ ) that causes the mispredictions, but the semantic analysis of the determiners G&S adopt.

information state is a set of possibilities, i.e., a set of world-assignment pairs. Discourse referents are interpreted as fixed variables of the assignments. The definition of the order  $<_P$  comparing the extensions of the question-predicate is extended to the case of possibilities by adding the condition that the assignments have to be identical to make possibilities comparable.<sup>27</sup> Dynamic exhaustive interpretation is then defined as a context change function that selects minimal possibilities instead of worlds.

### 3.5.1. DEFINITION. (Dynamic Exhaustive Interpretation)

Let  $A$  be an answer given to a question with question-predicate  $P$  in context  $\sigma$ . We define the exhaustive interpretation  $exh_{dyn}^\sigma(A, P)$  of  $A$  with respect to  $P$  and  $\sigma$  as follows:

$$exh_{dyn}^\sigma(A, P) \equiv \{i \in \sigma[A] \mid \neg \exists i' \in \sigma[A] : i' <_P i\}.$$

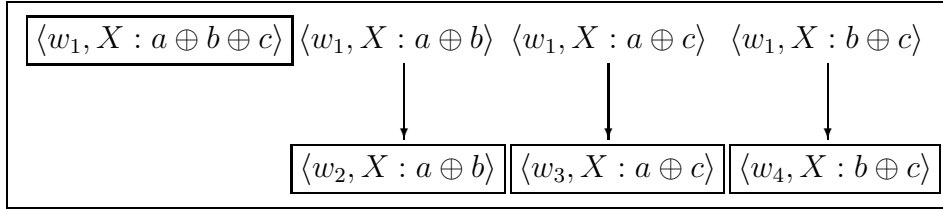
How does this straightforward extension of  $exh_{std}$  to dynamic semantics solve the problems discussed above? The crucial point is that the interpretation of introduced discourse referents becomes a fixed property of our possibilities. These variables can no longer be varied freely when the extension of the question-predicate is minimized. That makes it more difficult for possibilities to be minimal.

This will become much clearer after discussing some examples. First, let us consider the answers (33c), *Students*, and (33d), *Most students*. It has been argued that the determiners occurring in these answers belong to different classes. While the first (together with *A man*, *Some<sub>1</sub> girls*, *Five girls* and *At least five girls*) contains a weak determiner, the determiner of the second (together with *all ducks*, *most students*, and *some<sub>2</sub> girls*) is strong. Adopting a standard assumption of dynamic semantics (e.g. Kamp & Reyle, 1993), we treat only the latter type of NPs as two-place generalized quantifiers. NPs with weak determiners, in contrast, do not denote generalized quantifiers and directly introduce discourse referents. For anaphoric reference to strong quantifiers, discourse referents have to be constructed afterwards from the intersection of nucleus and restrictor. It turns out that if we adopt this treatment of weak and strong quantifiers the new function  $exh_{dyn}$  can account for the differences in the exhaustive interpretations.

Assume an information state:  $\sigma = \{i_1, i_2, \dots, i_8\}$ , where  $i_k = \langle w_k, g_k \rangle$ . In the worlds of all possibilities we have the same interpretation for *students*, the set  $\{a \oplus b, a \oplus c, b \oplus c, a \oplus b \oplus c\}$ . For the interpretation of *passed the examination* we assume:  $[P](w_1) = \{a, b, c, a \oplus b, \dots\}$ ,  $[P](w_2) = \{a, b, a \oplus b\}$ ,  $\dots$ ,  $[P](w_4) = \{b, c, b \oplus c\}$ ,  $\dots$ ,  $[P](w_8) = \emptyset$  - we simply take every possible distributive set given the three atoms  $\{a, b, c\}$ . Hence, predicate  $P$  is assumed to be distributive. Furthermore, notice that only in  $w_1$  but not in  $w_2$ ,  $w_3$ , and  $w_4$  it is true that all

<sup>27</sup>Thus, we redefine  $\langle w, g \rangle <_P \langle v, h \rangle$  iff<sub>def</sub>  $g = h$  and  $w <_P v$ .

students passed the examination. First, we calculate the exhaustive interpretation of answer (33c). After updating with the semantic meaning of the sentence *Students passed the examination*,  $\exists X : S(X) \wedge P(X)$ ,<sup>28</sup> we end up with an information state  $\sigma'$  containing successors of the possibilities  $i_1, i_2, i_3$ , and  $i_4$  whose variable assignments now are defined for  $X$ <sup>29</sup>, the newly introduced discourse referent.<sup>30</sup> In  $\sigma'$  there will be a possibility for every possible mapping of  $X$  to a group of students that passed the examination in one of the worlds  $w_1, w_2, w_3$ , and  $w_4$ . So  $\sigma'$  contains, for instance, the possibility  $\langle w_1, X : a \oplus b \rangle$ , because the object  $a \oplus b$  is in the extension of  $P$  in  $w_1$ . However,  $\langle w_2, X : a \oplus b \oplus c \rangle$  will not be an element of  $\sigma'$ . Given the assignment  $X : a \oplus b \oplus c$ , answer (33c) would not be true in  $w_2$ . The following tableau lists all possibilities in  $\sigma'$  plus the way they are ordered by  $<_P$ , where  $\langle \dots \rangle_1 \rightarrow \langle \dots \rangle_2$  means that the second possibility is  $P$ -smaller than the first.



To determine the exhaustive interpretation of answer (33c) we collect the minimal elements of this ordering (marked by a box in the picture) and obtain the set:  $\{\langle w_1, X : a \oplus b \oplus c \rangle, \langle w_2, X : a \oplus b \rangle, \langle w_3, X : a \oplus c \rangle, \langle w_4, X : b \oplus c \rangle\}$ . This interpretation still allows for the possibility that *all* students passed the examination - even though it would be excluded that anybody other than students passed the examination. The reason is that after updating  $\sigma$  with  $exh_{dyn}(\exists X : S(X) \wedge P(X), P)$  there is still a possibility that takes the world to be  $w_1$ : the possibility  $\langle w_1, X \rightarrow a \oplus b \oplus c \rangle$ . And this is so because there will be no possibility in  $\sigma[\exists X : S(X) \wedge P(X)]$  where the extension of  $P$  is smaller than in  $w_1$  and which still maps  $X$  to  $a \oplus b \oplus c$ . Such a possibility would not make the answer true. Hence, we correctly predict no *at most* inference for the answer (33c).<sup>31</sup>

However, doing the same calculation with *Most students*, the example (33d), will lead to a different result. Because strong determiners do not immediately introduce discourse referents, we obtain as the semantic meaning of the answer in context  $\sigma$  the set  $\{i_1, i_2, i_3, i_4\}$  (in all possibilities of  $\sigma$  it is true that most

<sup>28</sup>We take a standard approach to the dynamic meaning of  $\exists$  and interpret it as introducing a new discourse referent for the variable it binds.

<sup>29</sup>This suggests that we adopt a particular perspective on dynamic semantics where new discourse referents extend assignment functions. However, eliminative approaches to dynamic semantics work as well.

<sup>30</sup>The others are excluded by the truth conditions of the answer.

<sup>31</sup>Notice, by the way, that after exhaustive interpretation, if we refer back to the newly introduced discourse referent, we are talking about all students that passed the examination. This is on a par with intuition.



students passed). But  $i_2, i_3, i_4$  are all  $<_P$ -smaller than  $i_1$ . Thus, after exhaustive interpretation we end up with a new information state containing only  $i_2, i_3$  and  $i_4$ . The possibility that all students passed the examination is excluded.

Dynamic semantics also helps to account for the difference in exhaustive interpretation of (33a) *Three students passed* and (33b) *At least three students passed* (or answers like *Three or more students passed*). Within dynamic frameworks (e.g. Kamp & Reyle, 1993) it is standard to represent (33a) as  $\exists X : S(X) \wedge \text{card}(X) = 3 \wedge P(X)$ . This formula has the same *at least three* truth conditions that we obtain with the classical generalized quantifier interpretation of numerals. In particular, this sentence is true if four students passed, because then there is still a set of three students that passed. Thus, from a truth-conditional perspective we could have represented the semantic meaning of (33a) as well by  $\exists X : S(X) \wedge \text{card}(X) \geq 3 \wedge P(X)$ . Dynamically, however, the two formulas are not equivalent: the former introduces discourse referents that denote groups of exactly three individuals, while the groups introduced by the latter formula might be larger. As a consequence, if we apply  $exh_{dyn}$ , the former formula gets the *exactly three* reading, while the latter does not. This suggests that the former one correctly represents (33a), while the latter formula is the natural representation of answer (33b). And indeed, that was proposed by Kadmon (1985) (for related, but still somewhat different reasons). Hence, adopting Kadmon's analysis of the two determiners *three* and *at least three* allows us to account for their different behavior under exhaustive interpretation.<sup>32</sup>

To sum up the discussion in this section so far: the behavior of determiners is not a problem that forces us to give up the circumscription account for exhaustive interpretation or to propose that certain determiners have to come with special cancellation properties with respect to this mode of interpretation. It suffices to make the description sensitive to dynamic information.<sup>33</sup>

<sup>32</sup>Kamp & Reyle (1993), in fact, would not represent a sentence like (33b) by  $\exists X : \text{card}(X) \geq 3 \wedge S(X) \wedge P(X)$ , but rather by  $\exists X : \text{card}(X) \geq 3 \wedge X = \lambda y[S(y) \wedge P(y)]$ . For our purposes, however, this does not matter. They still predict that (33b) directly introduces a discourse referent and that is all we need for our analysis to go through.

<sup>33</sup>One may argue that free focus is generally interpreted exhaustively. However, certain examples suggest that in so-called topic-focus constructions, or sentences with a hat-contour, the focal-part should not be read exhaustively, even if it is used as an immediate response to a question. Consider (ii) and (iii) as answers to question (i).

- (i) What did the boys eat?
- (ii) [Some boys]<sub>T</sub> ate [broccoli]<sub>F</sub>.
- (iii) [One boy]<sub>T</sub> ate [broccoli]<sub>F</sub>.

If we would interpret *broccoli* exhaustively, and *some boys* or *one boy* as the generalized quantifier *at least some/one boy(s)*, it would mean for (iii) that for all alternatives  $x$  distinct to broccoli, the sentence *(At least) one boy ate x* has to be false. But this gives the wrong result that from (iii) we can conclude that *none* of the boys ate anything other than broccoli. As it turns out, also this problem disappears once we adopt a dynamic perspective. We interpret (iii), for

In the last part of this section we will discuss how dynamic information may also help to solve another part of the functionality problem of  $exh_{GS}$ . Remember example (28) *John or Mary or both*. The application of  $exh_{GS}$  to this answer (the same hold for  $exh_{std}$ ) excludes the last disjunct, hence, predicts that *either John or Mary is the only one who passed the examination*. But even though this answer can be interpreted exhaustively (implying that nobody besides John or Mary passed the examination) the possibility that both of them passed should not be excluded. Similarly, the sentence *John owns 3 or 5 cars* is on G&S's analysis (and by  $exh_{std}$  as well) falsely predicted to mean that John owns exactly 3 cars (the question is *How many cars does John own?* and we assume an *at least* interpretation of numerals). What we would like to end up with, however, is the prediction that John owns either exactly 3 cars, or exactly 5.

Intuitively, what both operations  $exh_{GS}$  and  $exh_{std}$  miss seems to be that in exhaustively interpreting an answer we are not allowed to exclude any possibility explicitly mentioned in the answer.<sup>34</sup> We can account for this using exactly the same strategy as for the closely related problem concerning determiners. One can simply propose that while *John or Mary passed the examination* and *John or Mary or both passed the examination* have the same truth conditions, their dynamic semantic meanings are, again, different. Maria Aloni (2003), for instance, argues for independent reasons that the first sentence should be represented by something like  $\exists q : \vee q \wedge (q = \wedge P(j) \vee q = \wedge P(m))$ , where  $q$  is a propositional variable and  $\vee$  and  $\wedge$  have their usual Montagovian meanings. Notice that this formula has the same truth conditions as the standard representation of the sentence:  $P(j) \vee P(m)$ . Following Aloni's lead, we should then, of course, represent *John or Mary or both passed the examination* by  $\exists q : \vee q \wedge (q = \wedge P(j) \vee q = \wedge P(m) \vee q = \wedge (P(j) \wedge P(m)))$ , which also gives rise to the same truth conditions. Still, with a dynamic interpretation of the existential quantifier the dynamic semantic meanings of the two formulas differ, because the latter allows for a verifying world-assignment pair where the assignment maps  $q$  to the proposition that both

---

instance, as  $exh_{dyn}^\sigma(\exists X[Boy(X) \wedge card(X) = 1 \wedge Ate(X, Broccoli)], \lambda y.Ate(X, y))$ . Sentence (iii) is now interpreted as stating that one boy ate broccoli, and that this one boy has eaten nothing else. We correctly predict that it is still possible that non-members of the denotation of the discourse referent  $X$  ate something other than broccoli, e.g. beans. Thus, we predict that examples (ii) and (iii) do not provide good arguments against an exhaustive interpretation of free focus.

<sup>34</sup>This was also the basic idea behind Gazdar's (1979) solution for this problem. He was not addressing the exhaustive interpretation of answers but analyzed the exclusive interpretation of *or* as scalar implicature. To account for the cancellation of this implicature in a sentence like *John or Mary or both passed the examination* Gazdar proposes that a disjunctive sentence additionally triggers the clausal implicatures that each of its disjuncts is considered possible. If the clausal and scalar implicatures of a sentence contradict each other – as is the case in the example at hand – clausal implicatures overrule scalar ones.

John and Mary passed the examination, while the former formula does not.<sup>35</sup> In almost exactly the same way as for the examples (33a) and (33b), this difference in dynamic semantic meaning has the effect that the two formulas give rise to different exhaustive interpretations: the former,  $exh_{dyn}^\sigma(\exists q : \forall q \wedge (q = \wedge P(j) \vee q = \wedge P(m)), P)$ , allows only for possibilities (and thus worlds) in which either only John or only Mary passed the examination; the latter,  $exh_{dyn}^\sigma(\exists q : \forall q \wedge (q = \wedge P(j) \vee q = \wedge P(m) \vee q = \wedge (P(j) \wedge P(m))), P)$ , allows for possibilities where both passed the examination. In a similar manner we can account for the exactly-reading of *John owns 3 or 5 cars*, if we represent it by  $\exists q : \forall q \wedge (q = \wedge [John\ owns\ 3\ cars] \vee q = \wedge [John\ owns\ 5\ cars])$ .

### 3.6 Exhaustivity and relevance

One problem of G&S's approach that our operation  $exh_{dyn}$  still inherits is that it cannot account for the contextual dependence of exhaustive interpretation we have observed in domain restricted exhaustive interpretations, the scalar readings, the mention-some readings, and differences in the fine-grainedness of the interpretation (see section 3.2.2). The crucial observations made when discussing these readings were that (i) in all these cases exhaustive interpretation was not substituted by some other interpretation but simply weakened<sup>36</sup>, and (ii) this weakening can be characterized as follows: inferences of the strong fine-grained mention-all reading of exhaustive interpretation disappear if they are commonly known in the context of utterance to be *irrelevant* for the questioner. This leads us to adopt the following strategy towards these readings: we extend our definition of exhaustive interpretation by making it dependent on what counts as relevant for the questioner. As it turns out, we can then correctly describe the intended variation in the strength of exhaustive interpretation.

Let us start with trying to understand what it means to be relevant information for the questioner and how it may play a role for the exhaustive interpretation of answers. If somebody poses a question, she is (normally) in need of certain information. A simple standard way to describe this information is by a set  $DP$

---

<sup>35</sup>Philippe Schlenker (p.c.) came up with a direct 'anaphoric' argument for why sentences of the form *A or B* and *A or B or both* indeed should have different dynamic semantic meanings:

- (i) We'll invite John or Bill, and *he*'ll have a good time.
- (ii) \*We'll invite John or Bill or both, and *he*'ll have a good time.
- (iii) We'll invite John or Bill or both, and *they*'ll have a good time.

These sentences suggest that the first conjunct of (i), for instance, should be represented by  $\exists x : Inv(x) \wedge (x = j \vee x = b)$  rather than by  $\exists q : \forall q \wedge (q = \wedge Inv(j) \vee q = \wedge Inv(b))$ . But this does not make any difference for our explanation.

<sup>36</sup>Thus, in contrast to other analyses we propose for mention-some readings that exhaustivity is not absent in these cases, but that it does not do anything.

of propositions.<sup>37</sup> For the questioner it is relevant to know which of these propositions actually hold. The semantic meaning  $Q$  of a question is also standardly described as a set of propositions, the appropriate, complete, or resolving answers to the question (see Hamblin 1973, Karttunen 1977, G&S 1984). It seems rational to assume that for reasons of efficiency there might be a difference between the information asked for explicitly by the questioner and the information needed, described by  $DP$ . For instance, assume that Ann is interested in who of John, Mary, and Peter passed the examination.<sup>38</sup> The question directly corresponding to this  $DP$  is *Who of John, Mary and Peter passed the examination?*. But it is arguably better for Ann to ask *Who passed the examination?*. A complete answer to this question would provide her with more information than she needs, but that does not bother her. However, the second question is shorter and thus spares her effort.

If it is commonly known what counts as relevant for the questioner, it would be reasonable for the answerer Bob to take this information into account as well and exhaustively specify only this part of the syntactic question-predicate that is relevant. Instead of listing all individuals that passed the examination, he only mentions whom of John, Mary, and Peter did. This spares *him* effort. Then, of course, a rational hearer will respect this factor as well when interpreting Bob's utterance and will not conclude from the answer *John* that John was the only individual that passed the examination, but rather that he was the only one of John, Mary, and Peter who did so. And this is exactly what seems to be going on in the case of domain restricted exhaustive interpretation.

Before we can come to a general formalization of this relevance-dependence of exhaustive interpretation, one further question has to be addressed: Does relevance already affect the interpretation of the question (thus, does Ann's question *Who passed the examination?* semantically mean *Who of John, Mary and Peter passed the examination?*) or is relevance independent contextual information? In the first case the description of exhaustive interpretation we have given can easily be made sensitive to relevance by proposing that the operation does not work on the syntactic question-predicate but rather on the predicate that the question is really about. The only thing that we have to do is to clarify how this predicate can be calculated given the semantic meaning of the question. If, however, the semantic meaning of the question is not affected by what is known about  $DP$ , then we cannot use this shortcut and have to incorporate relevance as a fourth argument into our definition of exhaustive interpretation.

This chapter is not about the semantics of questions. Therefore, we will not make a decision on this subject and shortly sketch how one can proceed in both cases distinguished above.

---

<sup>37</sup>This is a simplification of the notion of a *decision problem*.

<sup>38</sup>Thus,  $DP$  in this case is the set  $\{\{v \in W \mid [P](w) \cap \{j, m, p\} = [P](v) \cap \{j, m, p\}\} \mid w \in W\}$ .

### 3.6.1 The indirect approach

There are certain arguments that speak in favor of taking the semantic meaning of questions to depend on relevance. For instance, it seems that this factor also influences the interpretation of embedded questions. An extensive discussion of the pros and cons on this issue can be found in van Rooij (2003), followed by an approach to the meaning of questions that takes relevance into account. Also according to this approach the meaning of a question is a set of propositions – but now their semantics is underspecified with respect to contextual relevance. We will adopt this proposal here.

Given this position towards the semantics of questions we can make  $exh_{std}$ <sup>39</sup> dependent on relevance simply by manipulating its arguments. We propose that it does not apply to the syntactic predicate of the question asked but to that predicate whose extension the questioner is really interested in. We define it to be a minimal property  $\mathcal{X}$  such that knowing the extension of  $\mathcal{X}$  would resolve  $Q$ , whereby  $Q$  is the semantic meaning of the question asked. We say that  $\mathcal{X}$  is at least as minimal as  $\mathcal{Y}$ ,  $\mathcal{X} \subseteq \mathcal{Y}$ , iff  $\forall v : \mathcal{X}(v) \subseteq \mathcal{Y}(v)$ . Following G&S, we take *knowing*  $\mathcal{X}$  to mean knowing which of the following propositions is true:  $Q_{\mathcal{X}} = \{\{v \in W | \mathcal{X}(v) = \mathcal{X}(w)\} | w \in W\}$ . Thus, someone knows  $\mathcal{X}$  if she can specify for every object whether it has property  $\mathcal{X}$  or not. *Knowing the extension of  $\mathcal{X}$  resolves  $Q$*  is now understood as the following relation between  $Q_{\mathcal{X}}$  and  $Q$ :  $\forall q \in Q \exists q' \in Q_{\mathcal{X}} : q' \subseteq q$ , or informally, knowing the extension of  $\mathcal{X}$  has to imply knowing which elements of  $Q$  are true.

With this definition of the property the question is about we can correctly account for the different readings of exhaustive interpretation distinguished in section 3.2.2. For instance, if the whole extension of the syntactic predicate  $P$  of the question is relevant, van Rooij (2003) predicts that the meaning of the question is exactly such a partition  $Q_{[P]}$  as defined above. In this case, no smaller property  $\mathcal{X}$  than  $[P]$  itself will exist such that  $Q_{\mathcal{X}}$  solves  $Q_{[P]}$ . Hence  $\mathcal{X} = [P]$  and the speaker will by exhaustively specifying  $\mathcal{X}$  specify  $[P]$ . Thus, we predict a mention-all reading.

How to account for exhaustive interpretation when domain restriction or the level of required granularity is at issue is straightforward. For a degree-question like (26) *How far can you jump?*, for instance, the question-predicate can in some contexts range over meters, rather than centimeters. We therefore address scalar readings next. Remember the poker game example, (25) *What cards did you have?*. Again there exists a uniquely determined minimal  $\mathcal{X}$ , i.e., the  $\mathcal{X}$  that contains in every world those and only those cards that contribute to the value of the syntactic question-predicate  $P$  in terms of the card game. Exhaustifying this set tells us that the speaker had no additional cards that would increase the value of the cards she mentioned explicitly. Because in every world  $w$   $\mathcal{X}(w)$  is a

---

<sup>39</sup>To avoid unnecessary complications, we continue with  $exh_{std}$  in this section. However, the changes that will be proposed for this operation can be easily applied to  $exh_{dyn}$  as well.

subset of  $[P](w)$ , she may have had other, irrelevant, cards. About them, nothing can be inferred.

Finally, the mention-some case. Here, intuitively, the questioner does not care about what she learns about the extension of the syntactic predicate  $P$ , as long as she learns for one thing in its extension that it has property  $[P]$ .  $Q$  is predicted by van Rooij (2003) to be the set  $\{\{w \in W \mid d \in [P](w)\} \mid d \in D\}$ . Any predicate that in each world  $w$  applies to exactly one object in  $[P](w)$  and nothing else will qualify as a property  $\mathcal{X}$  that  $Q$  is about. In this case,  $\mathcal{X}$  is not uniquely defined. But for the interpreter the choice does not matter. In any case one learns from the answer that some subset of the extension of  $P$  consists exactly of the things mentioned in the answer - nothing more and nothing less. But that's exactly what one wants for an answer as in (23).

### 3.6.2 The direct approach

Assume that the semantic meaning of the question does not depend on relevance. How, then, can relevance influence the exhaustive interpretation of answers? A solution to this problem that still keeps the principal setting of our approach the same is to propose that the order  $\leq_P$ , on which the selection of minimal models is based, depends on relevance. In some sense this is what we have done in the indirect approach as well. We proposed that the predicate, or property, used in  $exh_{std}$  should be one that is partly defined in terms of relevance. Because the order  $<_P$  compares worlds with respect to the extension they assign to this predicate, the order becomes sensitive to relevance as well. Now, we have to change the definition of the order to make it directly dependent on some measure of relevance, not just on  $P$ .

We say that a world  $w_1$  is more minimal than a world  $w_2$ ,  $w_1 <_P^{rel} w_2$ , if (everything else being equal) the proposition claiming that all the objects in  $[P](w_1)$  indeed have property  $[P]$  is less relevant than the proposition saying the same for world  $w_2$ .

#### 3.6.1. DEFINITION. (The relevance order)

$$w_1 <_P^{rel} w_2 \text{ iff}_{\text{def}} \lambda w. [P](w_1) \subseteq [P](w) <_R \lambda w. [P](w_2) \subseteq [P](w)$$

By substituting this new order in the definition of  $exh_{std}$  we obtain a relevance dependent description of exhaustive interpretation.

#### 3.6.2. DEFINITION. (Relevance Exhaustive Interpretation)

Let  $A$  be an answer given to a question with question-predicate  $P$  in context  $W$ . We define the exhaustive interpretation  $exh_{rel}^W(A, P)$  of  $A$  with respect to  $P$  and  $W$  as follows:

$$exh_{rel}^W(A, P) \equiv \{w \in [A]^W \mid \neg \exists v \in [A]^W : v <_P^{rel} w\}$$

This leaves us with the problem to define the order  $\leq_R$  comparing the relevance of propositions. Fortunately, a lot of work has been done on this topic in decision theory that we can make use of. The account we propose here simplifies this work quite a bit, for we do not need full-blooded decision theory for our concerns. Remember that we described the information a questioner is interested in by a set of propositions  $DP$ . The questioner wants to know which one of these propositions is true. We define the *utility value* of a proposition  $p$  by how much it helps to select one of these propositions in  $DP$  as true. Let  $\mathcal{P}(q|p)$  be the quotient  $\frac{\text{card}([q]^W)}{\text{card}([p]^W)}$ . Hence,  $\mathcal{P}$  is a simplified measure of the probability of  $q$  given that  $p$  is true.<sup>40</sup>

### 3.6.3. DEFINITION. (The utility value)

$$UV(p) = \max_{q \in DP} \mathcal{P}(q|p) - \max_{q \in DP} \mathcal{P}(q|W)$$

The order  $<_R$  on propositions can then be defined by comparing this utility value.<sup>41</sup>

Now everything is in place and we can start to test the proposal. Assume that the speaker wants to know what exactly the extension of the syntactic question-predicate  $P$  is.  $DP$  is then the set of propositions that exactly specifies the extension of  $P$ . But this means that  $v <_P^{rel} w$  iff, everything else being equal,  $[P](v) \subset [P](w)$  and, thus, iff  $v <_P w$ . Hence,  $exh_{rel}^W$  reduces to  $exh_{std}^W$  and we predict a mention-all reading.

Mention-some answers can be treated as well. As already discussed in section 3.2.2, answers get mention-some or non-exhaustive interpretations in cases where it is clear that the addressee only has to know for someone in the extension of  $P$  that she has property  $P$ . For (23) *Who has a light?*, for instance, it is normally enough for Ann to know someone who has a light. She just wants to know who to ask for lighting her cigarette. Let us discuss a concrete example. Assume that  $D = \{j, m\}$ ,  $W = \{w_1, w_2, w_3\}$ , and  $[P](w_1) = \{j\}$ ,  $[P](w_2) = \{m\}$ ,  $[P](w_3) = \{j, m\}$ . Given what we have said about the questioner,  $DP$  would in this situation be the set:  $\{\{w_1, w_3\}, \{w_2, w_3\}\}$ . To determine the order  $<_P^{rel}$  we first have to calculate the utility values of the propositions  $\lambda w. [P](w_i) \subseteq [P](w)$  for  $i = 1, 2, 3$ . It turns out that  $UV(\lambda w. [P](w_i) \subseteq [P](w)) =$

<sup>40</sup>It is perhaps useful to point out that if  $DP$  is a singleton set consisting only of a ‘goal’ proposition  $h$ , the utility value we assign to a proposition comes down – according to definition 3.6.3 – to the standard statistical notion of relevance and is also very similar to what Merin (1999) defines as the relevance of this proposition. The notion given in definition 3.6.3 has also been used by Parikh (1992) and van Rooij (2003) for linguistic purposes.

<sup>41</sup>Notice that in case the reverse of  $<_R$  is one-sided entailment (see van Rooij (2004) for a discussion under which circumstances this will be so), it will be the case that  $w_1 <_P^{rel} w_2$  exactly if  $[P](w_1) \subset [P](w_2)$ . Thus, in these circumstances  $w_1 <_P^{rel} w_2$  if and only if  $w_1 <_P w_2$ : the old ordering between worlds is a natural special case of our new ordering. It follows that in these cases  $exh_{rel}^W(A, P) = exh_{std}^W(A, P)$ .

$UV(\lambda w.[P](w_2) \subseteq [P](w)) = UV(\lambda w.[P](w_3) \subseteq [P](w))$ . The order collapses totally, i.e., it does not make any difference which proposition is given as answer. Hence,  $exh_{rel}^W(P(j), P) = [P(j)]^W$ : our exhaustification operator adds nothing to the semantic meaning of the answer.

The other effects of exhaustive interpretation observed in section 3.2.2 can be treated in terms of  $exh_{rel}^W(A, P)$  successfully as well. Consider the granularity effect, for instance. If Ann is known to be only interested in the amount of (full) *meters* that Bob can jump, a world  $u$  where he can jump 5.00 meters, a world  $v$  where he can jump 5.50 meters, and a world  $w$  where he can jump 5.80 meters are all equally relevant,  $u \approx_P^{rel} v \approx_P^{rel} w$ . It follows that by taking Bob's assertion *I can jump five meters* exhaustively, we predict that in this case Ann can conclude only that he cannot jump six meters, but not that he cannot jump five meters and 10 centimeters.

This ends our excursion to a relevance-dependent notion of exhaustive interpretation. In the rest of the chapter we will ignore this possible extension and continue with the basic version of our proposal.

### 3.7 Exhaustive interpretation as conversational implicature

As the reader may have noticed, some inferences we have analyzed under the heading of exhaustive interpretation have also often been explained as *conversational implicatures*, in particular as *scalar implicatures*. To give two examples, the exclusive interpretation of *or* in (18): *John or Mary*, and the inference that not all students passed the examination from the exhaustive interpretation of (33d) *Most students* (the question which (18) and (33d) address is again *Who passed the examination?*) are standard scalar implicatures.

It turns out that the description of exhaustive interpretation proposed in the previous sections also correctly predicts many other scalar implicatures. To give an example, it also generates for (34b) the 'scale' reversal inference (34c).

- (34) (a) Ann: In how many seconds can you run the 100 meters?  
 (b) Bob: I can run the 100 meters in 12 seconds.  
 (c) Bob cannot run the 100 meters in 11 seconds.

The reason for this 'scale' reversal is that in contrast to predicates like *Bob owns  $n$  children* and *Bob can jump  $n$  meters far*, the question-predicate of (34a) behaves monotone increasingly in numbers:<sup>42</sup> if  $n$  is in the extension of the predicate and  $m > n$ , then  $m$  is in its extension as well.

Particularly pleasing is the observation that the approach to exhaustive interpretation defended here can also account for the well-known problematic cases

---

<sup>42</sup>This has to be guaranteed by meaning postulates.



of implicatures of complex sentences. For instance, using *exh<sub>GS</sub>* or *exh<sub>std</sub>* one can derive for multiple disjunctions as in answer (35) the inference that only one of the disjuncts is true (hence, only one of John, Mary, and Peter passed the examination).

(35) John, Mary, or Peter.

For example (36) we correctly predict that the interpreter can infer that John ate either only the apples or only some but not all of the pears.

(36) Ann: What did John eat?

Bob: John ate the apples or some of the pears.

Given these observations it is not very surprising that at different places in the literature it has been suggested that exhaustive interpretation can be explained as a pragmatic phenomenon using Grice's theory of conversational implicatures (see, for instance, Harnish 1976, G&S 1984). The central problem of such an approach is that there is no thoroughly satisfying formalization of Grice's theory and, hence, no precise description of the conversational implicatures an utterance comes with. But without such a rigorous description we cannot say whether Grice's theory indeed does account for some enrichment of semantic meaning. In particular, we cannot make such a claim for the exhaustive interpretation of answers. Thus, before we can see whether Grice's theory can be used to explain exhaustive interpretation, we first need to formally describe at least parts of the conversational implicatures an utterance comes with.

In Chapter 2 a new formalization of the Gricean reasoning leading to scalar implicatures has been proposed. We will follow van Rooij & Schulz (2004) in adapting this approach to the formal situation at hand but also add some small improvements.

The following Gricean principle has – in different forms – often been taken to be responsible for scalar implicatures. It combines Grice's first subclause of the maxim of quantity with the maxims of quality and relevance.

THE GRICEAN PRINCIPLE

In uttering *A* a rational and cooperative speaker makes a maximally relevant claim given her knowledge.

In the special case we are interested in here, where the utterance given is an answer to some previously asked question, the principle comes down to saying that the speaker will not withhold information from the audience that would help to resolve the question she is answering – she provides *a*<sup>43</sup> best (i.e. most relevant) answer she can, given her knowledge.

<sup>43</sup>There may be more than one optimum.

Our goal is to formalize the inferences an interpreter can derive if she takes the speaker of some sentence  $A$  to obey this principle. The solution proposed in Schulz (2005) and van Rooij & Schulz (2004) is closely related to McCarthy's predicate circumscription and makes essential use of ideas developed by Halpern & Moses (1984) on the concept of *only knowing*, generalized by van der Hoek et al. (1999, 2000). We describe the possibilities where the speaker obeys the principle as those where she knows the sentence  $A$  she uttered to be true but knows as little as possible about the predicate in question besides what is semantically conveyed by her answer. Hence, as in the case of predicate circumscription, the enriched interpretation of answer  $A$  is described by selecting minimal models. Now, however, the selection takes place among those possibilities where the speaker *knows*  $A$ , and the order that determines minimality does not compare the extension of the question-predicate, but rather how much the speaker knows about this extension.

To formalize such an interpretation function, we have to refer to the knowledge state of the speaker. We will adopt a standard modal logical way of modeling knowledge. Let  $W$  be a set of models/possible worlds of our language.<sup>44</sup> We add to  $W$  an accessibility relation  $R$  that connects every element  $w$  of  $W$  with a subset  $R(w)$  of  $W$ . This subset contains all worlds that are consistent with what the speaker knows in  $w$ . Then we can say that sentence  $\mathbf{K}A$ , *the speaker knows*  $A$ , is true in  $w$  (with respect to  $W$  and  $R$ ) if  $A$  is true in every world in  $R(w)$ . Because we want to model knowledge we demand that  $w$  is an element of  $R(w)$ . In this way we warrant that if the speaker knows  $A$ , the sentence is true in  $w$ .

Now, we can define an interpretation function that gives us besides the semantic meaning also the conversational implicatures due to the Gricean Principle. Assume that  $\preceq_{P,A}$  is the order that compares how much the speaker who uttered  $A$  knows about the question-predicate  $P$ .

### 3.7.1. DEFINITION. (Interpreting according to the Gricean Principle)

Let  $A$  be an answer given to a question with question-predicate  $P$  in context  $C = \langle W, R \rangle$ . We define the pragmatic interpretation  $grice^C(A, P)$  of  $A$  with respect to  $P$  and  $C$  as follows:

$$grice^C(A, P) =_{\text{def}} \{w \in [\mathbf{K}A]^C \mid \forall w' \in [\mathbf{K}A]^C : w \preceq_{P,A} w'\}$$

Of course, this definition will only be of use if we can also give an explicit definition of the order  $\preceq_{P,A}$ , and hence, describe what it means that in one possibility the speaker knows more about the extension of the question-predicate than in another. But when is this the case? Informally, what we want to express is that a speaker has more knowledge about  $P$  if she knows of more individuals that they have property  $P$ . Thus, we say that  $w_1 \preceq_{P,A} w_2$  if for every world  $v_2$  considered possible by the speaker in  $w_2$  (i.e.  $v_2 \in R(w_2)$ ), she distinguishes some possibility  $v_1$  in  $R(w_1)$  where the extension of  $P$  is smaller than or equal to the

<sup>44</sup>We allow multiple occurrences of the same interpretation function of the language in  $W$ .

extension of  $P$  in  $v_2$ .<sup>45</sup> But wait! It may be the case that the speaker makes in her utterance a claim about the extension of  $P$  which depends on some other facts. For instance, she may answer *If they asked the same questions as last year then Peter passed the examination* to the question *Who passed the examination?* Of course, in this case we expect a speaker that obeys the Gricean Principle also to tell us whether – as far as she knows – they asked the same questions as last year. Therefore, we define the order as follows:<sup>46</sup>

**3.7.2. DEFINITION.** Given a context  $C = \langle W < R \rangle$  we define for  $v_1, v_2 \in W$

$$\begin{aligned} v_1 \leq_{P,A}^* v_2 &\text{ iff}_{\text{def}} \quad \begin{aligned} &1. [P](v_1) \subseteq [P](v_2) \text{ and} \\ &2. \text{ for all non-logical vocabulary } \theta \text{ occurring in } A \\ &\quad \text{besides } P: [\theta](v_1) = [\theta](v_2); \end{aligned} \\ v_1 \equiv_{P,A}^* v_2 &\text{ iff}_{\text{def}} \quad v_1 \leq_{P,A}^* v_2 \text{ and } v_2 \leq_{P,A}^* v_1. \end{aligned}$$

**3.7.3. DEFINITION.** (Comparing relevant knowledge)

Given a context  $C = \langle W < R \rangle$  we define for  $w_1, w_2 \in W$

$$\begin{aligned} w_1 \preceq_{P,A} w_2 &\text{ iff}_{\text{def}} \quad \forall v_2 \in R(w_2) \exists v_1 \in R(w_1) : v_1 \leq_{P,A}^* v_2, \\ w_1 \cong_{P,A} w_2 &\text{ iff}_{\text{def}} \quad w_1 \preceq_{P,A} w_2 \ \& \ w_2 \preceq_{P,A} w_1. \end{aligned}$$

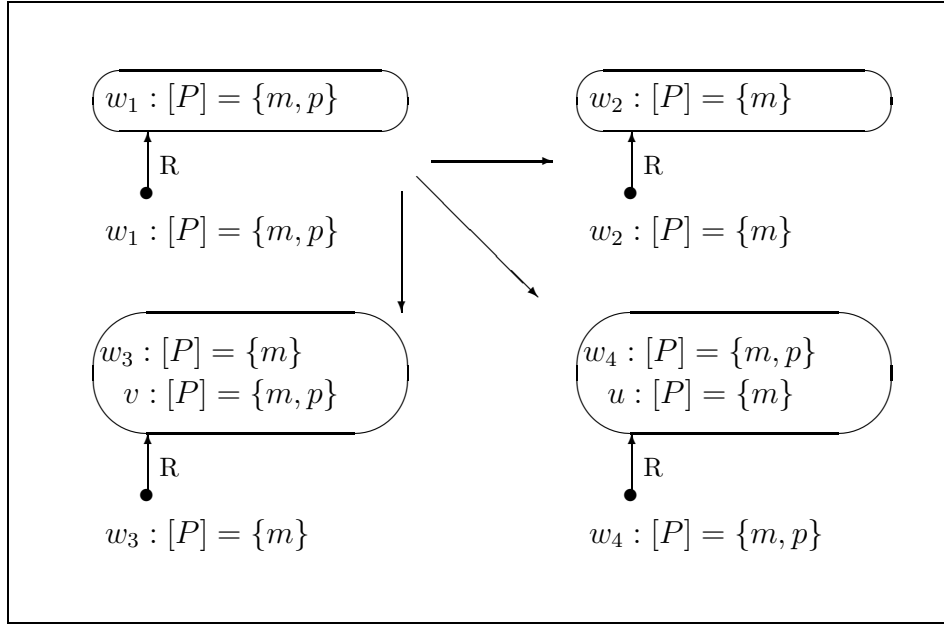
Now that we have with *grice* at least a partial description of the conversational implicatures an utterance comes with, we can see whether the part of Grice's theory we have formalized can explain the exhaustive interpretation of answers. Unfortunately, it turns out that this is not the case. To illustrate the problem, let us calculate what *grice* predicts for example (37).

- (37) Ann: Who passed the examination?  
 Bob: Mary.

Hence, let us determine  $\text{grice}^C(P(m), P)$ . We choose a model where for every individual there exists a unique name. To make things even simpler, we assume that in context  $C$  there are only four different worlds:  $w_1, w_2, w_3$ , and  $w_4$  with  $R$  and  $P$  defined as given in figure 3.1.<sup>47</sup>

<sup>45</sup>Some readers may notice that in this way we do not respect knowledge the speaker might have about some individuals not having property  $[P]$ . We would like to have some kind of motivation for why this information should not be taken into account, but until now we do not have a convincing explanation.

<sup>46</sup> $\leq_{P,A}^*$  is stronger than the order  $\leq_P$  that we have used so far. If one would substitute the latter in the definition of  $\preceq_{P,A}$ , then *grice* would minimize the knowledge of the speaker about the extension of *all* non-logical vocabulary, which is inadequate for our purposes. In van Rooij & Schulz (2004) an even stronger order was used. There, condition 2 of definition 3.7.2 was dropped and non-logical vocabulary besides  $P$  did not play any role for the order. Then, however, one misses for the answer *If they asked the same questions as last year then Peter*

Figure 3.1: The model for  $grice^C(P(m), P)$ 

What we would like to predict in such a situation is that all other individuals (in our example there is only one other individual: Peter (p)) did not pass the examination. To calculate  $grice^C(P(m), P)$  according to definition 3.7.1, the first thing we have to do is to select those worlds  $w$  in  $W$  where the speaker knows that  $P(m)$  is true. It turns out that this is the case for all elements of  $W$ . In a second step we select among those the possibilities where Bob knows least about the question-predicate  $P$ . The order tells us that the speaker knows more in  $w_1$  than in  $w_2, w_3$ , and  $w_4$ , and that in the latter three worlds he knows equally much. Hence:  $grice^C(P(m), P) = \{w_2, w_3, w_4\}$ . A closer look reveals that this interpretation allows the interpreter to derive from Bob's answer the conversational implicature that he does not know that Peter passed the examination (i.e.  $grice^C(P(m), P) \models \neg \mathbf{K}P(p)$ ). But we are not able to derive the desired inference that Peter, in fact, did not pass the examination. Hence, we have to conclude that the Gricean Principle, at least in the formalization given above, cannot explain exhaustive interpretation.

Actually, many students of conversational implicatures will find this a rather

---

*passed the examination* the intuitive inference that the speaker does not know whether they asked the same questions as last year.

<sup>47</sup>Possible worlds are represented by points. Arrows annotated with  $R$  lead from a world  $w$  to the knowledge state  $R(w)$  of the speaker in  $w$ . The arrows in the middle of the figure symbolize the ordering relation  $\preceq_{P,A}$ . Notice that in this example the worlds  $w_2$  and  $w_3$ , for instance, differ, because in  $w_2$  the speaker has a more definite opinion about the extension of  $P$  than in  $w_3$ .

pleasing result. It has often been argued in the literature that the conversational implicatures due to the Gricean Principle<sup>48</sup> should be generated primarily with the weak epistemic force we predict (see, among others, Soames 1982, Leech 1983, Horn 1989, Matsumoto 1995, and Green 1995). Hence, the conversational implicature of Bob's answer is indeed claimed to be that he does not know for people other than Mary that they passed the examination. Only in contexts where the speaker is assumed/believed to be competent/an authority on the subject matter under discussion, these authors propose, one can derive the stronger inference that what the speaker does not know to hold indeed does not hold (hence, in the example the desired inference that Peter did not pass the examination).

For our approach this would mean that we should be able to obtain the exhaustive interpretation by calculating *grice* with respect to the set *C* of contexts where the speaker Bob is competent/an authority on the question she is answering. However, in van Rooij & Schulz (2004) it is shown that this will not lead to an adequate description of the exhaustive interpretation of answers (or their scalar implicatures). The problem is that some sentences that can be interpreted exhaustively (or give rise to scalar implicatures) cannot stem from a speaker that is at the same time competent/an authority and obeys the Gricean Principle. An example is the answer Bob gives in (18), here repeated as (38).

- (38) Ann: Who passed the examination?  
 Bob: John or Mary.

In the present chapter we have proposed to analyze the often observed exclusive interpretation of *or* as due to exhaustive interpretation, and at many places in the literature the inference that not both disjuncts are true at the same time has been claimed to be a scalar implicature. However, the approach sketched above will not predict any conversational implicatures for such disjunctive answers. The reason is that a competent speaker should know which of the disjuncts is true and, if obeying the Gricean Principle, should have given this information. The fact that Bob nevertheless did not do so shows that he either is not competent or has disobeyed the principle. In neither case is the exclusive interpretation of *or* predicted.

To overcome this problem but nevertheless stay faithful to the intuition that competence/authority plays a decisive role for the derivation of exhaustivity effects/scalar implicatures, van Rooij & Schulz (2004) propose to *maximize* the competence of the speaker when interpreting answers. However, it is only maximized in so far as this is consistent with taking the speaker to obey the Gricean Principle.<sup>49</sup>

---

<sup>48</sup>In particular, conversational implicatures due to the first subclause of the maxim of quantity.

<sup>49</sup>In this respect, our analysis bears resemblance to Gazdar's (1979) proposal that clausal implicatures with weak epistemic force can cancel scalar ones that have strong epistemic force,

There is an obvious way to extend the function *grice* such that it follows this idea. We introduce a second order on the set of possibilities that compares the competence of the speaker in different possibilities (a possibility is higher in the order if the speaker is more competent). Then we select maximal elements with respect to this order – but now only among those possibilities where the speaker obeys the Gricean Principle, i.e., among the elements in  $\text{grice}^C(A, P)$ .

Let  $\sqsubseteq_{P,A}$  be the order that compares competence. The interpretation function defined below tells an interpreter what she can infer if she takes the speaker, first, to obey the Gricean Principle, and, second, to be as competent with respect to the question she answers as is consistent with the first assumption.

**3.7.4. DEFINITION.** (Adding Competence to the Gricean Principle)

Let  $A$  be an answer given to a question with question-predicate  $P$  in context  $C = \langle W, R \rangle$ . We define the pragmatic interpretation  $\text{eps}^C(A, P)$  of  $A$  with respect to  $P$  and  $C$  as follows:

$$\begin{aligned} \text{eps}^C(A, P) &=_{\text{def}} \{w \in \text{grice}^C(A, P) \mid \forall w' \in \text{grice}^C(A, P) : w \not\sqsubseteq_{P,A} w'\} \\ &= \{w \in [\mathbf{K}A]^C \mid \forall w' \in [\mathbf{K}A]^C : \\ &\quad w \preceq_{P,A} w' \wedge (w \cong_{P,A} w' \rightarrow w \not\sqsubseteq_{P,A} w')\}. \end{aligned}$$

Again, to make this definition useful we have to define the order  $\sqsubseteq_{P,A}$  properly. We propose that in a world  $w_2$  the speaker is at least as competent as in world  $w_1$  if in  $w_1$  the speaker considers as least as many extensions possible for question-predicate  $P$  as in  $w_2$ .<sup>50</sup>

**3.7.5. DEFINITION.** (Comparing competence)

Given a context  $C = \langle W, R \rangle$  we define for  $w_1, w_2 \in W$

$$w_1 \sqsubseteq_{P,A} w_2 \text{ iff}_{\text{def}} \forall v_2 \in R(w_2) : \exists v_1 \in R(w_1) : v_2 \equiv_{P,A}^* v_1.$$

To illustrate the working of the new strengthened interpretation function *eps*, let us reconsider our example (37) Ann: Who passed the examination? Bob: Mary. We calculate  $\text{eps}^C(P(m), P)$ , where  $W$  is defined as in figure 3.1. Remember that we want to obtain that Bob's answer implies that Peter did not pass the examination. Because we already know that  $\text{grice}^C(A, P) = \{w_2, w_3, w_4\}$ , the only thing that still has to be done is to select among the possibilities

---

and with Sauerland's (2004) method of strengthening implicatures with weak epistemic force. Our approach is based on essentially the same ideas as Spector's (2003) Gricean justification of exhaustive interpretation. In contrast to all these analyses, however, ours is more general, fully model-theoretic, and based on standard methods of non-monotonic reasoning.

<sup>50</sup>Here, again, we slightly deviate from the approach in van Rooij & Schulz (2004). There it is proposed to compare only what the speaker knows about objects that do not have property  $[P]$ . Because of the way that *grice* is defined, it does not make any difference for *eps* which of the two definitions is chosen.

in this set those where according to  $\sqsubseteq_{P,A}$  the speaker is maximally competent. Unsurprisingly,  $w_2$  is the unique  $\sqsubseteq_{P,A}$ -maximum in  $\{w_2, w_3, w_4\}$ . Hence,  $\text{eps}^W(P(m), P) = \{w_2\}$ . But, as figure 3.1 shows, in  $w_2$  the desired conclusion  $\neg P(p)$  holds! So, for this example the new interpretation function combining Gricean reasoning with a principle of maximizing competence predicts correctly.

Of course, we would like to establish that  $\text{eps}$  can account for the exhaustive interpretation of answers in some generality. At least it would be pleasing if  $\text{eps}$  does not perform worse in describing exhaustive interpretation than does  $\text{exh}_{std}$ . It turns out that both notions are indeed closely related.<sup>51</sup>

**3.7.6. FACT.** If  $A$  and  $\phi$  do not contain modal operators and  $C = \langle W, R \rangle$  is chosen such that there is no previous information in the context then

$$\text{eps}^C(A, P) \models \phi \text{ iff } \text{exh}_{std}^C(A, P) \models \phi.$$

Hence, if the answer given does not contain modal operators and there is no information in the context, then both interpretation functions predict the same modal-free inferences. Until now, all examples discussed in this chapter were of this kind. Thus, all the pleasing predictions made by  $\text{exh}_{std}$  are inherited by  $\text{eps}$ .

However, our richer modal analysis allows us, additionally, to describe exhaustivity effects that could not be accounted for in terms of  $\text{exh}_{std}$ . One kind of example are sentences that contain modal expressions, like belief attributions or possibility statements as in the answer *John and perhaps also Mary*. This advantage of the Gricean derivation of exhaustive interpretation given above is explicitly discussed in van Rooij & Schulz (2004) and we will not repeat that discussion here. In this section we only want to indicate (as also discussed in the paper mentioned) how the modal approach can help us to solve a last problem of G&S's (1984) approach that has not been addressed so far: the exhaustive interpretation of negative answers.

Remember our discussion at the end of section 3.3:  $\text{exh}_{GS}$  predicts wrongly that the answer *Not John* to question *Who passed the examination?* means that *nobody* passed the examination. The same prediction is made for all (other) cases in which the question predicate occurs under negation in the answer. To solve this problem, von Stechow & Zimmermann (1984) propose to modify G&S's exhaustivity operator by selecting in these cases not the *minimal* extensions of the predicate, but rather the *maximal* ones. In terms of our framework this means that now the speaker gives not the exhaustive extension of question-predicate  $P$ ,

---

<sup>51</sup>The extension of the definition of  $\text{exh}_{std}$  to context  $C = \langle W, R \rangle$  is straightforward. The proof of fact 3.7.6 goes very much along the same lines as the proof of a similar claim given in van Rooij & Schulz (2004). In the way the orders are defined here one needs the additional assumption that for all  $w \in [A]^C$  there is some  $w' \in \text{exh}_{std}^C(A, P)$  such that  $w' \leq_{P,A} w$ . Spector (2003) proves a closely related fact.

but of its *complement*,  $\bar{P}$ , instead. Thus, in case  $P$  occurs negatively in  $A$ , we should not look for  $exh_{std}^W(A, P)$  but rather for  $exh_{std}^W(A, \bar{P})$ . Then, the answer *Not John*, for instance, would be interpreted as implying that, except for John, everybody passed the examination. According to most of our informants, however, this kind of exhaustive interpretation is the exception, rather than the rule. They report instead that negative answers give rise to the conclusion that the semantic meaning of the answer is the only information the speaker has about the question-predicate. Interestingly enough, the same intuition is reported also for other answers, for instance, if the speaker uses special intonation or responds *Well/As far as I know, Peter*.<sup>52</sup>

A very welcome side-effect of the Gricean explanation given to exhaustive interpretation in this section is that we can correctly describe this interpretation when we only apply the function *grice*, hence, take the speaker to obey the Gricean Principle, but not maximize her competence. This suggests the following explanation for the non-exhaustive interpretation of the answers discussed above. The speaker is always taken to fulfill the Gricean principle and, hence, *grice* is applied to the answer. We normally also take the answerer to be competent<sup>53</sup> and, hence, apply *eps*. However, the answerer can cancel this additional assumption by either mentioning that she is not competent or simply deviating from the standard form of answering a question (by using negation, special intonation, etc.). In this way we can correctly predict the weakening of exhaustive interpretation to ‘limited-competence’ inferences for such answers.

### 3.8 Conclusion and outlook

In this chapter we did two things. First – and this was the central goal of the work presented – we propose a description of the exhaustive interpretation of answers. The main concept this description builds on is that of interpretation in minimal models, which we took from AI-research.<sup>54</sup> It constitutes the fundament of our formalization of exhaustive interpretation and holds the whole chapter together. The second backbone of our description is dynamic semantics. It provides us with the semantic framework in which we embedded minimal interpretation. And finally, we use standard conceptions of relevance to bring communicational

---

<sup>52</sup>See also footnote 23.

<sup>53</sup>This seems to be a natural default assumption, given that only in such situations it makes perfect sense to ask a question to a certain addressee.

<sup>54</sup>We only know of one (other) attempt to use circumscription for (some of) the data we discuss in this chapter: by Wainer in his dissertation (1991). When applying circumscription to utterances directly, he came across some of the same problems that we discussed for G&S’s proposal. For this reason he opts, in the end, for a second description in which stipulated abnormality predicates are circumscribed. One of the main goals of this chapter was to show that the direct approach without additional abnormality predicates can be pushed much further than Wainer assumed.



interests of the agents into play. Brought together, these three independent lines of research allow us to account for many observations on the phenomenon of exhaustive interpretation.

In the last section of the chapter we have gone beyond the primary goal to provide an adequate description of exhaustive interpretation. We used a proposal made in Schulz (2005) and van Rooij & Schulz (2004) to provide a pragmatic explanation for this rule of interpretation. Exhaustive interpretation is explained as based on the assumptions that first, the speaker obeys the Gricean Principle and, second, that she is competent on the question she answers (as far as this is consistent with the first assumption). We propose a formalization of these assumptions and the reasoning based on them that can be shown to perform as least as well in describing exhaustive interpretation as does  $exh_{std}$ . In fact, it turns out that this pragmatic explanation can account for a certain contextual weakening of exhaustive interpretation that none of our operations  $exh_{std}$ ,  $exh_{dyn}$  or  $exh_{rel}$  could deal with.

This part of our work allows us to answer a question that shadowed us through the whole chapter: what is the relation between exhaustive interpretation and conversational implicatures. According to us, exhaustive interpretation refers to a class of conversational implicatures, among them many scalar implicatures. It is the result of a Gricean-like reasoning about rational behavior of cooperative speakers. This explains why so often in the literature the notions *exhaustive interpretation* and *scalar implicature* are used to describe the same observation.

In the sections 3.5, 3.6, and 3.7 we have presented three different extensions of our basic description  $exh_{std}$  of exhaustive interpretation:  $exh_{dyn}$  was based on a dynamic approach to semantics,  $exh_{rel}$  took a contextual parameter of relevance into account, and  $eps$ , finally, modeled exhaustive interpretation as a consequence of a Gricean-like reasoning pattern. All of them addressed certain shortcomings of our initial account in terms of  $exh_{std}$ , but none of them overcomes all of them. The ultimate goal should be to combine all these extensions into one uniform description. We did not present the account in this way, because it would have made the chapter much less readable. The interaction between the different extensions raises many additional questions that have to be addressed carefully. For instance, one can easily define  $eps$  based on a dynamic semantics with the aim of giving a Gricean motivation for  $exh_{dyn}$  as an extension of our justification for  $exh_{std}$ . But then one has to deal with questions such as in how far should information the speaker has about discourse referents involved in the answer be taken into account when comparing her knowledge? For some answers to these questions we will be able to present a dynamic version of fact 3.7.6, for others not. To keep these complications out of the already quite demanding discussion of the chapter, we decided to split up our approach in different units and present them separately. However, this should not make the reader lose sight of the composite form of our proposal.

Above we said that we take scalar implicatures to be a subclass of the inferences of exhaustive interpretation. On the other hand, at the beginning of the chapter we introduced exhaustive interpretation as the (normal) interpretation of answers. On the face of it this would mean that we predict scalar implicatures to be restricted to answers – what some of our readers may think a rather dangerous claim. But, first, the fact that exhaustive interpretation as discussed here was restricted to answers to overt questions does not necessarily mean that it occurs only in these contexts. This restriction was forced upon us mainly because we did not have sufficient empirical data to support a general statement about the contexts in which exhaustive interpretation occurs. Furthermore, one of the central issues in the recent literature on scalar implicatures is the context-dependence of these inferences. In particular, it has been claimed that questions can play an important role for the presence of scalar implicatures (see, for instance, Hirschberg 1985 and van Kuppevelt 1996). Further research on this subject has to clarify in which contexts we do observe scalar implicatures, and whether they coincide with the contexts of exhaustive interpretation.

Another interesting question for further research is whether the given formalization of the Gricean Principle can be extended to a general implementation of Grice's maxims of conversation. Consider, for instance, the second subclause of the maxim of quantity. This subclause is taken to be the driving force behind another class of pragmatic inferences: those to the most stereotypical interpretation. For instance, that we normally interpret *John killed the sheriff* as meaning that John murdered the sheriff in a stereotypical way, i.e. by knife or pistol, is often explained with reference to this maxim. Inferences to the stereotype/normal case (called *I*-implicatures by Atlas & Levinson 1981, and *R*-implicatures by Horn 1984) are often analyzed as being in some sense opposite to scalar implicatures.<sup>55</sup> Against this background it is interesting to observe that the minimal model approach can be used naturally to account for the latter inferences as well.<sup>56</sup> The only thing that we have to change is how we instantiate the ordering. In this case it is not predicate minimization that counts – of which relevance minimization is a natural extension – but rather minimization of *normality* (or maximization of *plausibility* or *expectedness*). Thus, now we have to assume that  $v \prec_A w$  iff  $v$  is a less surprising  $A$ -world than  $w$  is, and the interpretation of  $A$  w.r.t.  $\prec_A$ ,  $\{w \in [A]^W \mid \neg \exists v \in [A]^W : v \prec_A w\}$ , results then just in the set of most plausible worlds that verify  $A$ . In the future we would like to see to what extent this formalization can account for the wide range of *I* or *R* implicatures described by

---

<sup>55</sup>The intuition being that while in case of scalar implicatures some stronger claim is excluded, in case of inference to the stereotype some stronger claim is assumed to hold.

<sup>56</sup>In fact, these are the inferences non-monotonic reasoning was originally made for. See, for instance, McCarthy's motivation for introducing Predicate Circumscription as briefly discussed in section 3.4.1

Atlas & Levinson and Horn as due to this maxim,<sup>57</sup> and how they interact with scalar implicatures.

Of course, the observation that both types of implicatures may be captured by very similar interpretation rules does not make them necessarily the same phenomenon. In AI there has been an intense debate on the interpretations that non-monotonic reasoning formalisms can receive. One of the distinctions made there seems to show up here again. We have described exhaustive interpretation as a rule of negation as failure in the message: from the fact that the speaker did not say  $p$  for a certain class of propositions  $p$ , the interpreter infers that  $\neg p$ . Already McCarthy (1986) mentioned such rules of language use as examples of circumscription in action. A similar rule may also govern the *I* or *R* implicatures: if the speaker did not mention that the situation is in a certain way abnormal, then the interpreter can conclude that it is normal. But here we do not have to take the detour via language use. It may also simply be the case that the interpreter concludes to the stereotypical interpretation because it is for her the normal state of affairs given the information she has (including the message of the speaker). Note that this is not an admissible interpretation of exhaustive readings: if we learn that Mary has property  $P$ , only in very exceptional cases will general knowledge about how the world normally is allow us to infer that John does not have property  $P$ .

Hence, in summary, while the inference of negation as failure inherent in exhaustive interpretation is most plausibly due to rule-governed conversational behavior (which may be conventional or not), the inference to the stereotypical interpretation does not need to be anchored in language use.

---

<sup>57</sup>In several papers, e.g. Asher & Lascarides (1998), a sophisticated method of non-monotonic reasoning is used to account for some of these inferences.

### 4.1 Introduction

The subject of the second part of the dissertation are English conditional sentences. More particularly, we will investigate the question, whether it is possible to give a compositional account to the meaning of English conditionals. Because conditionals are quite complex expressions it would be too bold an aim to investigate this question for all parts of the construction. We will therefore focus on one aspect that is particularly problematic: deriving compositionally the temporal properties of conditionals from the tense markings present in their form.

Approaches to the meaning of conditionals can be traced far back in the history of philosophical thinking. This has certainly to do with the close connection between conditionals and reasoning: conditionals are the natural language expression of reasoning from hypotheses to conclusions. A huge part of this literature looks on the meaning of conditionals at a very abstract level, taking them to be expressions of the form  $A \succ C$  where  $A$  and  $C$  are sentences and  $\succ$  a binary sentence connective expressing the conditional relation between those sentences. The way these expressions relate to English conditionals is left implicit. This is not very surprising, given that most of this work stems from philosophers and not from linguists. But for a linguist this abstract look on conditionals is not satisfying. For instance, a linguist is not satisfied with analyzing antecedent and consequent as primitive sentences (or maybe combinations of such primitive sentences connected by the standard logical operators *and*, *or*, and *not*). The extra structure present in natural language conditionals, like tense operators, modals, etc., matters for the meaning of conditional sentences. This thesis purports to answer some of the *linguistic* questions the semantics of English conditional sentences raises.

This is not the first linguistic work on the compositional semantics of conditionals. In fact, one can observe a recent growth of interest in the compositional semantics of conditionals (Iatridou 2000, Ippolito 2003, Kaufmann 2005, Asher

& McCready 2007 and others). The present work differs from other proposals that pursue a similar direction in how close it gets to the ultimate goal of a full compositional semantics for conditionals. It will distinguish contributions made by the modals, tenses, and the perfect, and it will deal at the same time with indicative as well as with subjunctive and counterfactual conditionals. Furthermore, the approach will make very specific and formally precise claims about the compositional semantics of English conditionals. We thereby hope to raise the level of detail of the discussion and to stimulate more specific empirical investigations into the meaning of conditionals.

Our main focus is on deriving the temporal properties of conditionals from the tense and aspect markings occurring in them. Why focus on this particular aspect of the compositional semantics of conditionals? The reason is that there is something strange going on with the interpretation of the tenses and the perfect in English conditionals. Their interpretation does not behave as one would expect. Consider, for instance, the interpretation of the simple past in conditionals. Sometimes it does behave as expected. The meaning of the conditional in (39) with an antecedent marked for the past tense can be paraphrased as *If at some past time during the morning of the utterance day Peter took the plane ....* Thus, the antecedent is analyzed as describing some event situated before the utterance time.

(39) If Peter took the plane (an hour ago), he will arrive in Frankfurt this evening.

But now consider (40). In this case the antecedent is about Peter's taking the plane in the future of the utterance time! Even without any temporal adverbial the antecedent cannot be interpreted as localizing the evaluation time of the antecedent in the past. Thus, surprisingly, in conditional sentences the past tense sometimes refers to the future.

(40) If Peter took the plane (in an hour), he would arrive in Frankfurt this evening.

Something similar is illustrated with example (41). Even if not obligatory in this case, the consequent of the conditional can be interpreted as describing an interview that will take place in the future. The central goal of Chapter 6 of this dissertation is to explain such and related puzzles concerning the interpretation of the tenses and the perfect in English conditionals.

(41) If Peter comes out smiling, the interview went well.

Before we come to the temporal properties of conditionals, we will first, in Chapter 5, start at the level of abstraction which is so omnipresent in the literature on conditionals, and discuss the meaning of a conditional  $A \succ B$  in a

timeless framework, ignoring to a great extent the compositional structure of antecedent and consequent. One reason is that a lot of interesting things have been said about the meaning of conditionals at this level and we want to take this work into account when developing a compositional approach to the meaning of these sentences. But the main motivation is to resolve certain open questions at this abstract level. They concern in particular the semantics of counterfactual conditionals. We will try to answer these questions before matters get complicated by the introduction of time into the model. We will then use these answers when developing the more complex, compositional, approach to the meaning of conditional sentences in Chapter 6.

## 4.2 Central ideas

### Changing the facts *versus* changing your beliefs

There are two central claims of the theory developed here that characterize the approach and place it in relation to other theories on the same subject. One of these central claims is that a systematic distinction has to be made between an *epistemic* and an *ontic* reading of conditionals. This distinction is present for all types of conditionals, indicative conditionals as well as subjunctive conditionals or counterfactuals. Even though some authors make similar claims (see, for instance Kaufmann, 2005), this ambiguity has often been ignored, because the two readings can only rarely be distinguished for one and the same conditional sentence. Either they are identical, or one (the epistemic reading) is only marginally available. This makes it also difficult to illustrate the ambiguity, but let's try. Consider the following example.

*Last night the duchess was murdered in her sleep. You are supposed to find the murderer. Soon after the investigations started the lab calls. They have found fingerprints of the butler all over the crime scene. You interrogate the butler and he confesses. At this state you believe that the butler did it, and that the gardener had nothing to do with it. Somewhat later the lab calls again. They have checked all the locks of the house. None is broken. There are only two persons besides the duchess that have keys for the house: the butler and the gardener. Now, you believe:*

*(42) If the butler had not killed her, the gardener would have.*

The theory that we will develop will predict that according to the dominant ontic reading the conditional (42) is false, but according to its marginal epistemic reading it is true. While most people agree on the existence of the first reading, some deny the possibility of the second for this example.

The two readings for conditional sentences will not be modeled – as has often been proposed (see, for instance, Kratzer 1979, 1981) – by letting conditionals refer to different modal bases, but by distinguishing two ways to update an information state with a sentence. These two update functions represent two different perspectives on how to act with language. The epistemic interpretation function is based on descriptive language use. It takes the sentence with which the information state of the interpreter is to be updated as providing information about the actual world. The ontic interpretation function assumes a prescriptive language use. It makes the sentence with which the information state is to be updated true in all possibilities of the information state (if possible). The distinction of different interpretation functions is, even though not unique, quite unusual in formal semantics. It appears to go against the central goal of classical formal semantics, which is to remove all the ambiguities in natural language. In the formal semantics developed here every expression has two interpretations. However, our approach does not predict a systematic ambiguity of English utterances. Because the two interpretation functions implement two different types of speech acts, it is the language's and the speaker's capacity to distinguish between these two speech acts that disambiguates the interpretation.

In this dissertation we restrict our attention to assertions, i.e. descriptive language use. That means that on the level of sentences the epistemic interpretation function always has to be applied. But we will also propose that there are some lexical items the epistemic interpretation of which can make reference to the ontic update function. Among these are the modals *will*, *would*, *may*, and *might*, as well as the sentence connective *if*. Because of this we also have to provide a description of the ontic interpretation function when dealing with assertive language use.

### The mood of English sentences

Another central claim of the theory of English conditionals developed here is that English assertive sentences are obligatorily marked for mood. We will propose that in contemporary English three moods have to be distinguished for assertions: an *indicative* mood, a *subjunctive* mood and a *counterfactual* mood. The mood gives information about how the content of a sentence relates to the information about the actual world already present in an information state. In particular, it helps to determine when the sentence gives information about a subordinate, hypothetical belief state. For instance, we will predict for example (43) that the subjunctive mood marked on the modal in the third sentence is responsible for this sentence being about some hypothetical context introduced in the second sentence and not about the actual world. Hence, the subjunctive is responsible for the sequence of sentences in (43) only having a reading that says that if the thing we are hearing snooping around next door is a wolf, it would eat you first and not as saying that the thing next door, whatever it is, would eat you first.

- (43) There is something snooping around next door. It might be a wolf. It would eat you first.

We will propose, furthermore, that the subjunctive and the counterfactual mood are marked in English using the simple past and the past perfect. Hence, according to the approach developed here, the form of the simple past and the perfect are ambiguous between a temporal/aspectual meaning and a mood meaning. This will explain why sentences like (44) are not about the past. In this case the past morpheme is interpreted as subjunctive mood.

- (44) If Peter took the plane (that leaves in an hour), he would arrive in Frankfurt this evening.

Introducing ambiguities into the lexicon is always unwanted. But, as we will argue in Chapter 6, in this case it is the best explanation for the missing past interpretation of examples like (44) at hand.

### 4.3 Terminological preliminaries

Before the real work starts we will first introduce some basic terminology, in particular make clear what we understand by English *conditional sentences* within this dissertation. We take such a sentence to consist of a main or matrix sentence, called the *consequent*, and a subordinate sentence starting with *if*, called the *antecedent*. Other sentences sometimes analyzed as conditionals, such as cases where the antecedent comes without *if*, but starts with the auxiliaries *were*, *should*, *might*, or *could*, are ignored, as well as cases where the antecedent is given in another form than as subordinate sentence, or is left implicit in the context. We will also not deal with conditionals the consequent of which starts with *then*.

We will follow an established praxis in English grammars and distinguish three different conditional constructions of English, although our terminology and definitions may differ. These three types are *indicative conditionals*, *would conditionals* and *would have conditionals*. We will refer to the latter two types together also as *subjunctive conditionals*. In the literature a mixture of semantic and syntactic criteria is often used to distinguish between these three types. This confusion can lead to fundamental misunderstandings about their meaning. Therefore, we will be very explicit on this point and base our definition only on syntactic properties of conditionals. In the following, we will define all three types of conditionals and add some further observations on each of the classes.

**Would conditionals.** *Would* conditionals are sometimes also called *non-past hypothetical conditionals* or *future less vivid conditionals*. Some authors also subsume them under *counterfactuals*. All these alternative notions are based on semantic criteria. Our criterion, given in the definition below is purely syntactic.



**4.3.1. DEFINITION.** *would* conditionals

A *would conditional* is a conditional sentence that contains as main finite verb in the consequent *would*, *could*, *might*, or *should*, not followed by a perfect. The antecedent stands in the simple past, not followed by a perfect.

We observe that in *would* conditionals the antecedent may contain modal verbs as well. Modals that may occur are *could* and *should*, and in exceptional cases also *would*. Various authors have observed that in the antecedent the modals cannot have non-root meanings.<sup>1</sup> Another thing to notice about *would* conditionals is that the finite verb in the antecedent can be *were*. *Were* is the past subjunctive of *be* and the last surviving past subjunctive form in English. Some examples for *would* conditionals are given below.

- (45) a. If Peter took the plane (in an hour), he would arrive in Frankfurt this evening.
- b. If I were you, I would leave him.
- c. If I won the lottery, I would buy a car.
- d. If I could ski, I would join you.

**Would have conditionals.** Other names for *would have* conditionals one can find in the literature are *past hypothetical conditionals* or *counterfactuals*.<sup>2</sup>

**4.3.2. DEFINITION.** *would have* conditionals

A *would have conditional* is a conditional sentence that contains as main finite verb in the consequent *would*, *could*, *might*, or *should*, followed by the perfect auxiliary *have*. The antecedent stands in the past perfect.

As in *would* conditionals the modals *could* and *should*, and in exceptional cases *would*, can occur in the antecedent followed by the perfect. Again, we add some examples for *would have* conditionals.

- (46) a. If I had won the lottery, I would have bought a car.
- b. If you had been in Paris next week, we could have met.

---

<sup>1</sup>That means that an epistemic or ontic/metaphysic reading of the modals is not possible. Instead the modal has to be interpreted as deontic or as referring to the abilities of some relevant person.

<sup>2</sup>Let us point out again that what other authors may mean with these notions can differ in details from what we defined to be a *would have* conditional.

**Indicative conditionals.** The name indicative conditionals is quite standard for the group of conditionals we refer to here. Sometimes, one also finds the notion *open conditional*. We define this type of conditional as not falling in one of the two first groups.

#### 4.3.3. DEFINITION. Indicative conditionals

An *indicative conditional* is a conditional sentence that contains as main finite verb in the consequent none of *would*, *could*, *might*, or *should*.

In the antecedent the finite verb can stand in the present or past tense. The perfect may be used. As modal *can* may occur followed by the infinitival main verb, in exceptional cases an occurrence of *will* is also possible. In most indicative conditionals containing a modal the finite verb in the consequent is *will*. Its place can also be taken by the modals *can*, *may*, *might*, *must*, *shall* or *pres(be) going to*. In contrast to the other two types of conditionals, the consequent of an indicative conditional may also be free of modals. In this case the finite verb in the consequent can be in the present or the past tense. If the bare simple present is used, then the conditional gets a habitual reading (see (47b)). This is not obligatory for past tense consequents.

- (47) a. If I win the lottery, I will buy a car.  
 b. If butter is heated, it melts.  
 c. If he leaves the interview smiling, it went well.  
 d. If the condition was not met, then the program flow skips past the statement in the <statementsX> element, at which point, another ELSEIF keyword or the ELSE, or END IF keywords are expected. (google example)  
 e. If no previous condition has been met within a blocked style If Selection Construct, and the optional ELSEIF keyword is encountered, the condition specified by the <conditionX> element is evaluated. (google example)  
 f. If democracy continues to struggle, then it has not arrived. (google example)

Even though the given classification covers most examples of conditional sentences, it is not exhaustive. Antecedents and consequents of different types can be mixed, as, for instance, the following example from google shows. The conditional (48) cannot be a subjunctive conditional, because the finite verb in the antecedent does not stand in the past tense. It can also not be an indicative conditional, because of the modal *would* in the consequent. In consequence, this conditional falls in none of the three groups specified above. It combines the antecedent of an indicative conditional with the consequent of a *would* conditional.

- (48) If HP has ironed out the issues, I would be the first to purchase this one when it becomes available. (google example)

There seem to be certain restrictions on possible combinations. For instance, sentences like the following, that combine antecedents of *would* conditionals with consequents of indicative conditionals, appear to be generally unacceptable.

- (49) If you were the richest man on earth, I will marry you.

A convincing theory of the semantics of English conditionals should be able to account for possible combinations like (48) and explain why certain combinations like (49) are out. The theory developed in this dissertation makes predictions on this point, but before we can test the theory in this respect, more empirical investigations of the issue are needed. Hopefully, this question will be studied in the future in more detail.

In the following we will often discuss temporal properties of conditionals and consider questions like the temporal reference of antecedent and consequent relative to the utterance time, to each other, etc. We will introduce some terminology here that allows us to make such statements in a clear, but fairly pre-theoretic, manner. For simple modal free tensed sentences we will distinguish the *evaluation time* of the sentence, which is roughly the time at which the eventuality described in the sentence is located. This time is restricted by the tense of the sentences and can further be restricted by temporal adverbials occurring in the sentence. The tense localizes the evaluation time relative to some *reference time*<sup>3</sup>, which is in most cases the *utterance time*. But, as we will see, it can also lie in the future of the utterance time. To illustrate the working of this terminology, for sentence (50) uttered by me at the moment I typed it the utterance time is 11:26 on March 20th, 2007, the reference time is this time as well, and the evaluation time is some time in the morning of March 20th, 2007.

- (50) It was snowing in Amsterdam this morning.

For simple sentences with modals we have to distinguish two evaluation times: the evaluation time of the modal phrase and the evaluation time of the phrase in the scope of the modal. Thus, for sentence (51) uttered at 11:34 on March 20th,

---

<sup>3</sup>We make a different use of the term *reference time* than does Reichenbach (1947) in his famous approach to the meaning of the English tenses. What Reichenbach means by reference time when talking about the past perfect is in our terminology the evaluation time of the perfect phrase in scope of the simple past of a past perfect construction. Our use of the notion *reference time* also differs from the use made by Kamp & Reyle (1993). They refer with it to what they call the ‘anaphoric dimension’ of tenses of natural language: tenses locate the evaluation time of the phrase in their scope relative to a contextually given ‘reference point’. Kamp & Reyle (1993) want to account this way for the temporal relations between sentences in discourse.

2007, the utterance time is 11:34 on March 20th, 2007, the reference time is 11:34 on March 20th, 2007, the evaluation time of the modal phrase is 11:34 on March 20th, 2007 and the evaluation time of the phrase in scope of the modal is one hour later.

(51) In an hour I will go for lunch.

We make a similar distinction for simple sentences with the perfect. Thus, for sentence (52) uttered at 11:36 on March 20th, 2007, the utterance time is 11:36 on March 20th, 2007, the reference time is 11:36 on March 20th, 2007, the evaluation time of the perfect phrase is 11:36 on March 20th, 2007 and the evaluation time of the phrase in the scope of the modal is some time in the past of 11:36 on March 20th, 2007.

(52) Simon has lost one of his gloves.

As a consequence, for modal perfect constructions three evaluation times have to be distinguished. If the modal scopes over the perfect, then these three times are (i) the evaluation time of the modal phrase, (ii) the evaluation time of the perfect phrase, and (iii) the evaluation time of the phrase in scope of the perfect.

Let us add a final side-mark. Neither in the discussion of the data nor in the formalization will we make a distinction between the evaluation time of some sentence and the temporal trace of the eventuality described by the sentence. Even though they cannot be identified in general, the difference is not directly relevant for the questions we try to answer in this work.

## 4.4 Caveat lector

Every research project has to be clear about its scope, the boundaries within which it is carried out. The resources of research are limited, certainly in the case of a dissertation project. Some issues, even though relevant for the topic of investigations cannot be dealt with. In this section we will set some boundaries for the research reported here.

In the section on terminological preliminaries it should have become clear that this research is about very specific conditional sentences: those with an explicit antecedent in the form of a subordinated sentence starting with the connective *if*. But there are also other ways in which the class of conditionals discussed here is restricted. We will not consider conditional sentences with a non-assertive consequent, as in (53a) or (53b). We will also exclude Austin conditionals (example (53c)) and anankastic conditionals (example (53d)). These conditionals raise a lot of issues of their own that we cannot deal with here. However, we hope that with the right semantics for questions, imperatives, etc. the approach presented in the following chapters can be easily extended to account for these conditionals.

- (53) a. If this has been discussed before then please stop me, but ... (conditional plea)
- b. If all Prophecy has been fulfilled, then isn't the Bible Irrelevant? (conditional question)
- c. If you are thirsty, there is beer in the fridge. (Austin conditionals)
- d. If you wanna go to Harlem, you have to take the A-train. (anankastic conditional)

Furthermore, we will not consider habitual readings of conditionals. That means, we will not deal with conditionals that make statements about general regularities like (54). Indicative conditionals without a modal in the consequent in particular favor habitual readings. For those the consequent of which stands in the present tense and is not marked for the perfect, it has even often been claimed to be the only possible reading. If this is correct, then we have nothing to say about sentences like (54).

- (54) If butter is heated, it melts.

There is also another way in which the scope of the theory developed here is limited, except for the class of conditional sentences it considers. The compositional analysis of conditional sentences that will be proposed is not complete. We will only analyze the structure of conditional sentences to the extent that the contribution of the tenses, the perfect and the modals can be distinguished. Predicate structures and reference to individuals will not be distinguished. Even more important, we will not deal with the aspectual classes of the verbal phrases in antecedent and consequent. Although we do propose a semantics for the tenses and the perfect, it will only concern their temporal properties, not their aspectual impact. This decision allows us to keep our model simple in that we do not have to introduce event semantics. However, because the topic of the present research concerns the temporal properties of conditionals, and aspectual questions are without doubt of relevance for these temporal properties, this is a limitation of the present work that has to be overcome in future work.

Finally, a more methodological caveat. People familiar with the classical literature on the semantics of conditionals may miss a study of the logical properties the theory developed here predicts for conditionals. The discussion in more traditional approaches is concerned with establishing the validity or invalidity of certain logical principles that are considered to characterize conditional reasoning. However, many of these issues, which play an important role in philosophical discussions, are orthogonal to the more linguistic questions we want to answer here. Thus, even though it is interesting and relevant to investigate how the present approach behaves with respect to these logical properties, this is not our priority here, and therefore has to await future work.

## Chapter 5

---

# The meaning of the conditional connective

### 5.1 Introduction

A large part of the extensive philosophical literature on the semantics of conditionals deals exclusively with the meaning of *would have* conditionals. Additionally, philosophers generally describe the semantics of these sentences at a very abstract level, ignoring the semantic impact of tense and modality markers etc. occurring in antecedent and consequent on the semantics. *Would have* conditionals are treated as constructions made up of a conditional operator  $\succ$  and two sentences representing antecedent and consequent. These sentences are taken to be, if not primitive, then combinations of primitive sentences using the standard connectives  $\wedge$ ,  $\vee$  and  $\neg$  and sometimes also  $\succ$  itself. As explained in the introduction, we want to extend this line of approach with a more serious consideration of the compositional structure of English conditional sentences, to deal in particular with their temporal properties. But this certainly does not mean that all the classical, abstract work on the meaning of *would have* conditionals is useless. In the ideal case we can take a description of the meaning of *would have* conditionals at this abstract level as a starting point and obtain a linguistically more adequate approach by the introduction of a more complex formal language and the addition of time to the model. The problem with this strategy is that there is not such a thing as *the* approach to the meaning of *would have* conditionals at the traditional level of abstraction. Instead there are many different proposals that all have been criticized for various reasons. There is, however, one approach that is particularly popular and has dominated the thinking on the meaning of *would have* conditionals through the last decades. This is the *similarity approach*

proposed by Stalnaker (1968) and Lewis (1973). According to this approach, a *would have* conditional is true if on those models for the antecedent that are most similar to the evaluation world the consequent is true as well. Unfortunately, this approach also comes with certain problems. Hence, we cannot take it unqualified as a starting point for our work. A central criticism is that the description of the similarity relation provided in the original work of Stalnaker and Lewis is too vague. The present chapter will address this problem and try to solve it at the traditional abstract level, before matters become complicated by the introduction of time into the model and a much more complex syntactic analysis of conditional sentences. The goal is first to come up with a convincing description of the semantic meaning of *would have* conditionals at the traditional, abstract level that can then be used as a starting point for the compositional approach developed in the following chapter.

The chapter is structured as follows. We will start by giving a short outline of the basic idea of the similarity approach. Afterwards we will discuss two types of observations brought forward to argue for a more restricted notion of similarity than what was originally proposed by Stalnaker and Lewis. In particular, the observations have been used to defend the popular idea that similarity is prominently similarity of the past. We will argue that while indeed there is evidence showing that a more restricted notion of similarity is needed, the conclusion that this restriction has to apply to some notion of pastness is not necessary. Furthermore, we will claim that such a purely temporal restriction of similarity is not appropriate to describe the meaning of *would have* conditionals.

We turn then to another proposal for how to restrict the similarity relation: *premise semantics*. Premise semantics combines the similarity approach with a different tradition in the history of approaches to the semantics of *would have* conditionals: *cotenability theories*. According to premise semantics similarity has to be defined in terms of a certain set of facts of the evaluation world. In the simplest case a world is said to be more similar to the evaluation world, the more of these facts it makes true. Theories of premise semantics can differ in the facts they take to be relevant for similarity and how exactly the impact of these facts on the similarity relation is described. We will focus on one recent proposal in this framework (Veltman 2005) and show how it solves some of the problematic examples for the similarity approach. However, we will also see that it is not able to provide the right restrictions for similarity in general.

After this we will turn our attention to an approach to the meaning of *would have* conditionals that – at least on first view – diverges from the similarity paradigm. This is the proposal made in Pearl (2000), according to which *would have* conditionals are interpreted as executing hypothetical surgeries on causal dependencies. We will argue, that while this approach can account for many traditionally hard examples for *would have* conditionals, it makes the wrong assumption that all of these conditionals are based on causal dependencies.

The proposal of Pearl and premise semantics are then combined and turned into a new approach to the meaning of *would have* conditionals. We will argue that two readings for *would have* conditionals have to be distinguished. One reading – the *ontic* reading – can be described by an adapted version of Pearl’s theory. The second reading – the *epistemic* reading – is based on belief revision. Both readings will be formalized as instantiations of the similarity approach, more particularly, of premise semantics. We will see that the new theory obtained this way allows us to overcome many of the shortcomings of the other approaches discussed in this chapter.

## 5.2 The similarity approach to conditionals

Since the early seventies the dominant approach to the meaning of *would have* conditionals is the similarity approach. This approach is based on the influential work of Stalnaker (1968) and Lewis (1973). The basic idea behind the similarity approach is very simple. In a possible world model we propose that a conditional with antecedent A and consequent C is true if on a certain selected set of worlds where the antecedent is true the consequent is true as well. Different types of conditionals may impose different restrictions on the relevant set of antecedent worlds. For indicative conditionals it has often been proposed that the set of antecedent worlds that are consistent with the epistemic state of some agent is relevant. For counterfactual conditionals, of course, this is not an option. Because of their counterfactuality there cannot be any world consistent with the beliefs of some agent where the antecedent is true. Thus, according to such a theory counterfactual conditionals would be trivially true. But we also cannot take in this case all worlds in which the antecedent is true as the relevant set on which the consequent has to be evaluated. Consider, for instance, the *would have* conditional (55). Standing outside on the street I say:

(55) If I had dropped this glass, it would have broken.

Intuitively, this sentence is true. If in order to evaluate this counterfactual we would take all counterfactual worlds into account where the antecedent is true, then among these worlds there would also be worlds where the street is not like my street made of stone, but just a sand road. In this case the glass might not have broken. Hence, if all antecedent worlds were considered, the counterfactual would come out as false. The idea of the similarity approach is that at least for counterfactuals, but maybe also for other types of conditionals, the antecedent worlds on which the consequent is tested are those<sup>1</sup> that are maximally similar

---

<sup>1</sup>According to Stalnaker (1968) there is only one single world making the antecedent true that is most similar to the evaluation world. But this particularity is not relevant for our discussion.



to the evaluation world. To come back to the example, because in the world where (55) is evaluated the street is made of stone, in the maximally similar worlds where I drop the glass the street is made of stone as well. Therefore, the conditional is evaluated to be true.

In their original work Stalnaker (1968) and Lewis (1973) leave the specific character of the similarity notion very unclear. Lewis advises us to understand similarity as *overall* similarity. This is vague, Lewis agrees. But he claims that the vagueness left by this description fits the obvious vagueness of the meaning of counterfactuals (Lewis 1973: 91). Later on he adopts the position of Stalnaker who claims that the similarity relation is semantically underspecified, but that the context pragmatically fills in the missing details. One problem for both positions is that the resulting theory for the meaning of conditionals is barely testable. Even stronger, many authors have argued that the proposed underspecification of the similarity relation is simply inadequate to capture the truth conditions of *would have* conditionals. They claim that if the similarity approach is correct, then there have to be some general semantic restrictions on what counts as similar. In the next section we will discuss two types of observations brought forth to support this claim and the particularly popular conclusion that similarity has to be restricted to or at least dominated by similarity of the past.

### 5.3 Similarity as similarity of the past

It has often been proposed in reaction to Stalnaker and Lewis' work that similarity should be specified as similarity of the past.<sup>2</sup> What exactly that means has been spelled out in very different ways, but basic to all these approaches is the idea that in some sense facts about what happened before a certain reference point count more for similarity than facts that are about the future of this reference point. A very famous proposal along these lines is Thomason and Gupta (1981). They propose that "past closeness predominates over future closeness" (p. 301). Some authors, for instance Arregui (2005), go even as far as claiming that *only* the past counts for similarity (where 'past' may mean different things in different approaches). Also Lewis himself proposed in later work (Lewis 1979) that there are certain restrictions on the similarity relation that make it behave asymmetrically with respect to the past and the future. But before we adopt such a restriction on the similarity relation we first have to ask what kind of (linguistic) evidence exists in favor of it. We will discuss two arguments that have been often

---

<sup>2</sup>In the previous section we said that in this chapter we will abstract away from issues of time. Now it might appear as if time makes its reappearance via the backdoor. Indeed, theories that follow these lines do need time to describe the meaning of conditionals even on the present abstract level. We will, however, argue that the similarity relation does not make any reference to time. Thus, the approach that will be adopted later will indeed not need a time-sensible model.

brought forth in the literature to support this claim. The first argument is the issue of backtracking *would have* conditionals. The second is what Lewis calls the *future similarity objection*.

### 5.3.1 Backtracking counterfactuals

The term *backtracking counterfactuals* was introduced by Lewis (1979). It refers to *would have* conditionals where the evaluation time of the phrase in the scope of the modal in the consequent lies temporally before the evaluation time of the antecedent. It has often been claimed that there are restrictions on the acceptability of backtracking counterfactuals, restrictions that are not shared by backtracking indicative conditionals. Take, for instance, the following two sentences.

(56) a. If he came out smiling the interview went well.

b. ??If he had come out smiling, the interview would have gone well.

Various authors observe that while the indicative conditional (56a) is fine, the counterfactual variant (56b) is not acceptable.<sup>3</sup> Observations like this have been taken to show that there is a special role for past to play in the semantics of *would have* conditionals, or, with respect to the similarity approach, that the past dominates the similarity relation (see, for instance, Frank 1997 and Arregui 2005). The reasoning is that backtracking counterfactuals like (56b) are unacceptable, because in the most similar worlds where the antecedent is true the past is unchanged – at least in those respects relevant for the truth of the consequent. Thus, is the conclusion, in some sense the past is not changed as easily as is the future in counterfactual reasoning. This is then proposed to originate in the higher relevance of the past for the similarity relation.

In this section we will review the different observations concerning backtracking counterfactuals made in the literature. After this we want to discuss the question whether we can conclude from these observations that similarity has to be restricted to or dominated by some notion of pastness.

A milestone in the literature on backtracking counterfactuals is Lewis (1979). Right at the beginning he observes “Seldom, if ever, can we find a clearly true counterfactual about how the past would be different if the present were somehow different. Such a counterfactual, unless clearly false, normally is not clear one way or the other.” (Lewis 1979: 455). Although this nearly seems to contradict this citation, according to Lewis backtracking is rare, but it is not impossible. It is

---

<sup>3</sup>These judgments refer to the reading of the two conditionals according to which the subject is smiling after leaving the interview. Sentence (56b) is fine in a reading where the interview takes place after the subject leaves somewhere smiling.

not quite clear, however, under what circumstances Lewis considers backtracking possible. In one passage he describes these contexts as those where extraordinary circumstance concerning time, causation etc. exist, as for instance “... in a time machine, or at the edge of a black hole, or before the big bang, ...” (Lewis 1979: 458]). On the other hand he gives an example for a backtracking counterfactual he considers acceptable that does not seem to be situated in a world with abnormal conditions concerning time and causation (Lewis 1979: 456, the example goes back to Downing 1959).

“Jim and Jack quarreled yesterday, and Jack is still hopping mad. We conclude that if Jim asked Jack for help today, Jack would not help him. But wait: Jim is a prideful fellow. He never would ask for help after such a quarrel;

(57) If Jim asked Jack for help today, there would have to have been no quarrel yesterday.

In that case Jack would be his usual generous self. So if Jim asked Jack for help today, Jack would help him after all. At this stage we may be persuaded (and rightly so, I think) that:

(58) If Jim asked Jack for help today, there would have been no quarrel yesterday.”

There seems to be some disagreement between linguists concerning the acceptability of this example. Frank (1997), for instance, takes Lewis to claim that (58) is not acceptable. But in most papers referring to Lewis adopt his judgment that (58) represents an acceptable backtracker in this context. We will adopt this position here. If, however, the example is fine, then we have to conclude that true backtracking counterfactuals do exist, and that they do not necessarily involve reasoning about abnormal conditions concerning causality or time.

In a side-mark Lewis notices an important peculiarity of backtracking counterfactuals that is also observed at other places in the literature. Very often, acceptable backtrackers have an additional modal *have to* in their consequent. Lewis does not consider its presence necessary but he thinks that (57) is more natural than (58). Lewis deliberately uses such a construction in (57) to “... lure you into a context that favors backtracking” (Lewis 1979: 458). Similar observations can be found, for instance, in Simon & Rescher (1966). In other papers it is claimed that *have to* does not simply improve on the acceptability or idiomaticity of backtracking counterfactual, but turns otherwise unacceptable instances into acceptable *would have* conditionals ( Frank 1997, Arregui 2005).

Even though there might be some disagreement concerning Lewis’ original example, it is quite generally accepted that some backtracking *would have* conditionals are acceptable, even without the presence of an additional *have to*. Some examples are given below.

- (59) a. If Clarissa were 30 now, she would have been born in 1966. (Frank 1997: 297))
- b. If he were a bachelor, he wouldn't have married. (Arregui 2005: 85)
- c. If Stevenson had been President in February 1953, he would have been elected in November 1952. (Bennett 1984: 79) (Stevenson lost the presidential elections to Eisenhower in November 1952)

An eye-catching difference in the discussion of these examples compared with Lewis' discussion of example (58) is that here the authors do not find it necessary to "lure" the reader into accepting the backtracking conditionals. Thus, it seems that there is a difference in acceptability of examples like (59a)-(59c) and backtrackers like (58). The question, then, is what is responsible for this difference. An interesting proposal is made by Arregui (2005). She claims that the distinguishing factor between both types of examples is that *would have* conditionals like (59a)-(59c) are based on non-contingent, logical/analytical laws. Indeed, there seems to be a systematic difference in the intuitive acceptability of backtrackers based on analytical/logical truths and backtrackers that are based on natural/causal laws. We therefore propose the following two generalizations of the observations on backtracking counterfactuals.<sup>4</sup>

- Generalization 1: Backtracking counterfactuals can straightforwardly be judged true, if the relation between antecedent and consequent is analytic/logically necessary.
- Generalization 2: Backtracking counterfactuals not based on such analytical/logical truths but on other types of generalizations are less acceptable, but not excluded. Such conditionals improve if *have to* is inserted in the consequent.

To further support the generalizations we show that when a non-analytical and non-logical law underlying counterfactual reasoning is turned into an analytical truth, the backtracking counterfactual clearly increases in acceptability. For illustration we use our former example (56b). In the context provided below the original causal relation between a successful examination and a happy face is turned into a convention and, thus, an analytical law. Now, the conditional (56b), here repeated as (60), is fine.

*The day of the final school examinations is approaching. Bill and his best friend Tom both have to meet the professors at the same day, but Bill's appointment is earlier than that of Tom. Tom would like to know*

---

<sup>4</sup>See related generalizations in Arregui (2005).

*whether everything went well with Bill before he has to enter the examination room. However, students that have already been examined are not allowed to talk to those still waiting. Therefore, they arrange that when leaving the building Bill will smile if his examination went well.*

*On the very day of the final examination, Tom and Sue are standing outside to school waiting for Bill's reappearance. Bill comes out looking rather displeased and walks away. Tom says to Sue:*

(60) If Bill had left smiling, the interview would have gone well.

Let us, for a moment, turn away from backtracking counterfactuals and consider *have to* insertion in *would have* conditionals in general. While *have to* always improves on the acceptability of backtracking counterfactuals, this is not the case for *would have* conditionals in general. We even observe that sometimes insertion of *have to* decreases the acceptability of a conditional (see the example given below, but make sure that you do not read the extra modal *have to* as root modal, expressing obligation).

*Sue and Tom had a serious fight. Sue needed a car and just took Tom's without telling him. Tom got very angry about it. Now he has taken the keys away from Sue and told her that he will never give her the car again. Sue complains about this to her friend Mary, but Mary responds: 'Why didn't you ask him for the car?' ...*

(61) a. If you had asked him, he would have given it to you.

b. ??If you had asked him, he would have to have given it to you.

Consideration of similar examples suggests that one has to distinguish between different readings of *would have* conditionals and that it depends on the choice of the reading whether *have to* insertion improves or weakens the acceptability of the conditional. This is illustrated with the next two examples. The same conditional is used in two very similar contexts. In the first case the extra modal *have to* is fine and even improves the acceptability of the conditional, in the second case *have to* insertion lessens the acceptability of the conditional.

#### Example 1

*Sue and Tom are back from their holidays. When they arrive in their flat, Sue tries to call her mother to tell her that everything went well, but it turns out that the phone has been disconnected. Sue asks Tom whether he has paid the bill. Tom insists that he did. Sue says: 'I don't believe you.'*

- (62) a. If you had payed the bill, the telephone would be working.  
 b. If you had payed the bill, the telephone would have to be working.

Example 2

*Sue and Tom are back from their holidays. They are both very hungry from the long drive home. When they arrive in their flat, Sue tries to call her mother to tell her that everything went well, but it turns out that the phone has been disconnected. Sue asks Tom whether he has payed the bill. Tom says: No, I didn't. You know that I couldn't pay the bill, because my paycheck was late. Sue says: 'Yes, that's true. But it is a pity'*

- (63) a. If you had payed the bill, the telephone would be working and we could have ordered a pizza.  
 b. ?If you had payed the bill, the telephone would have to be working and we could have ordered a pizza..

Intuitively, in either case the conditional expresses two different kinds of reasoning from antecedent to consequent. In the first example the conditional expresses what Sue would conclude, if she accepted the antecedent as part of her beliefs. This is essentially epistemic reasoning. In the second example the conditional is not about what Sue would have believed on learning that the antecedent is true, but about how the true antecedent would have changed the course of history. If our observations are correct, one could propose the following empirical generalizations.

- There are two readings of *would have* conditionals, an epistemic reading and a second reading, let's call it the ontic reading.
- The epistemic reading improves with *have to* insertion (or at least the acceptability of the conditional does not decrease). The ontic reading is not admissible when this modal occurs in the consequent of a conditional.
- From our earlier observation about the general improving effect of *have to* insertion in backtracking counterfactuals it follows that the epistemic reading allows for backtracking, but the ontic reading does not.

We will come back to these observations later.

Let us summarize our findings on backtracking. The picture you get when looking at the literature is far from clear. Not only do different papers report different intuitions for the same examples – even within one paper it is sometimes not clear what exactly the judgments are concerning particular examples. The lesson

we should learn from this is that we have to be very careful with deriving far-reaching conclusions about the semantics of *would have* conditionals from these shaky grounds. Before we can build theories about backtracking we need a far better empirical basis to start with. None of the papers reported on in this section has carried out serious empirical research, nor will we improve on them in this point. In view of the diversity of intuitions, there is an urgent need for such investigations.

Given that we cannot say much, what *can* we say about backtracking so far? Clearly, backtracking arguments are rare. The majority of *would have* conditionals does not locate the consequent before the antecedent, nor do they use backtracking reasoning to derive the antecedent. However, it would be false to claim that backtracking is in general not possible. Backtracking based on analytic/logical truths is uncontroversially acceptable and true. Also backward reasoning on natural/causal laws is not generally out or false. *Would have* conditionals based on such reasoning are acceptable at least in some circumstances to some (not necessarily few) speakers. The acceptability improves, when an additional *have to* is inserted after *would* in the consequent.

Let us finally say something about the relevance of backtracking counterfactuals for semantic theories of *would have* conditionals. It is important to realize that the observations made do not necessarily show that past plays a special role for the meaning of *would have* conditionals, more particularly, that similarity is systematically restricted to or dominated by the past. There are other possible explanations. Actually, most papers on the issue of backtracking seek the explanation rather in properties of the laws/generalizations underlying the conditional reasoning. Lewis suggests that the fact that earlier affairs are extremely over-determined<sup>5</sup> by later ones, but less so the other way around is responsible for the temporal asymmetry of counterfactual reasoning.<sup>6</sup> Frank (1997) proposes that it is the concept of historical necessity, or, as Lewis (1979) calls it, the asymmetry of open future and closed past, that is responsible. Arregui (2005) blames (besides temporal properties of *would have* conditionals) the different characteristics of the underlying laws for the restrictions on backtracking.

Except for the possibility of alternative explanation for the observations on backtracking, there is also a potential danger in making temporal conditions on

---

<sup>5</sup>A set of facts *S* determines a fact *p*, if *p* can be derived from *S* given the laws of nature.

<sup>6</sup>From this asymmetry of over-determination it follows that convergence to a world takes much more of a miracle (violation of natural laws) than diverging from it. Lewis proposes that the similarity relation is sensible to the size of miracles that take place in a world. Backward counterfactual reasoning is out, because the similarity relation allows for small miracles (divergence from *w*), if this enlarges the region of maximal overlap. Therefore, the antecedent is made true by a small miracle. Forward reasoning is fine, because it would take a big miracle to make all the consequences of the antecedent undone (convergence to *w*) and big miracles lead to very far-fetched worlds. For more details on this derivation of the asymmetry of counterfactual reasoning see Lewis (1979).

the similarity relation responsible for the observations. As we have seen, backtracking is in principle possible. Thus, at least to a certain extent the past changes under conditional reasoning. It is questionable whether temporal properties of the similarity relation alone can model the relevant type of changes in the past that is admitted.

### 5.3.2 The future similarity objection

We now come to the second argument brought forward to show that if the similarity approach is correct, then similarity is either restricted to or dominated by some notion of pastness. Lewis (1979) calls this argument the *future similarity objection*. The version of the future similarity objection Lewis cites in his paper (1979) stems from Fine (Fine 1975: 452).

“The counterfactual

- (64) If Nixon had pressed the button there would have been a nuclear holocaust.

is true or can be imagined to be so. Now suppose that there never will be a nuclear holocaust. Then that counterfactual is, on Lewis’ analysis, very likely false. For given any world in which the antecedent and consequent are both true it will be easy to imagine a closer world in which the antecedent is true and the consequent false. For we need only imagine a change that prevents the holocaust but that does not require such a great divergence from reality.”

According to the similarity approach, for the conditional (64) to be true the worlds most similar to the evaluation world where Nixon pressed the button have to be such that in them the nuclear holocaust does take place. However, or so Fine argues, in some of the most similar worlds no nuclear holocaust will take place. Think of worlds where, for instance, the button has been disconnected from the atom-rocket. Fine argues that some changes of reality that stop the nuclear holocaust from happening – like cutting the wire to disconnect the button – will, overall, change the course of events less than the nuclear holocaust itself. Therefore, a world where, for instance, the wire is cut, is more similar to the actual world than a world where the nuclear holocaust takes place. The obvious reaction to this criticism – and this is also the one given by Lewis (1979) – is that this argument refers to a notion of similarity that is not the one underlying counterfactual reasoning. But this raises the question about the character of the notion of similarity the meaning of *would have* conditionals is based on. It has to be such that disconnecting the circuit counts for more than the holocaust does. Fine (1975) and others have concluded that the reason for this difference in impact on the similarity relation is that disconnecting the circuit happens before Nixon



presses the button and the holocaust happens after this. Hence, they conclude that temporal properties of the involved events make a difference for similarity. More particularly, facts about the past count more for similarity than facts about the future.

We will argue that this is not convincing. First, notice that we cannot account for the acceptability of (64) just by proposing that worlds that differ from the actual world only at or after the temporal location of the eventuality described in the antecedent are more similar than worlds differing already before this time. It is possible that something happened to the mechanisms launching the rockets after Nixon pressed the button but before the nuclear holocaust takes place. Maybe someone cuts the wire exactly in this tiny span of time before the rockets are launched. To account for the intuitive correctness of (64) but also to respect this possibility, a proponent of the similarity-of-the-past idea could propose that differences between worlds count the more for similarity the further in the past they are. However, this will not work either, as the following example shows.<sup>7</sup>

*A farmer uses the following strategy to turn sheep into money. First he tries to sell a sheep to his brother. If he doesn't want it, it gets special feeding and some weeks later the farmer tries to sell it to the butcher. If the butcher doesn't want it, he gives it as a gift to the local zoo. One of the sheep is a particular favorite with his little son Tom. Tom doesn't know what became of Bertha, his favorite, because he was away for four weeks. The first thing he does after coming back is checking where Bertha is. He hears that his uncle bought her. Tom says that he is happy that he hasn't have to pay to visit her, because:*

(65) If my uncle hadn't bought her, she would have been a gift to the zoo.

Intuitively, this conditional is false in the given context.<sup>8</sup> However, the approach sketched above would predict the sentence to be true: the butcher had to buy or refrain from buying Bertha before she was offered to the zoo. Hence, if the restriction on similarity described above were in force, then a world where the butcher did not buy Bertha should be closer to the actual world, than a world where the butcher bought her. Therefore, in the most similar worlds making the antecedent of (65) true the consequent is true as well.

Maybe one can come up with other ideas how to restrict similarity purely by reference to temporal differences between worlds that can deal with Fine's example and the Bertha example as well. We cannot exclude this possibility

---

<sup>7</sup>This example is based on an example taken from Bennett (2003).

<sup>8</sup>The intuitions changes, if one assumes, for instance, that Tom knows that the butcher was not going to buy Bertha. But we assume that the context gives a complete description of the facts relevant for the evaluation of the example.

here. What we have seen is that it is clearly not an easy enterprise to find working temporal restrictions. There is, however, a very natural and simple alternative how to account for the example without reference to time. In an intuitive sense worlds where the wire is cut are much further off than worlds where the nuclear holocaust takes place. The holocaust – and all of its consequences – does not count for similarity on top of the fact that Nixon pushed the button, because they are *consequences* of this fact. We expect from Nixon’s pressing the button that the holocaust will take place, because we assume a law-like connection to hold between these two facts. Cutting the wire is not a natural consequence of pressing the button. Therefore it counts for similarity. Thus, to evaluate the conditional (64) we only compare worlds with respect of whether they agree with the evaluation world in that the wire is not cut, not with respect to whether the holocaust takes place or not. In consequence, the worlds Fine (1975) considers problematic come out as less similar than those worlds where the holocaust takes place. The conditional (64) is predicted to be true, as intended.

We conclude that there is a very intuitive alternative to the similarity of the past account of the Nixon example. One can – and we will claim later that one should – account for the truth of (64) by assuming that facts from which together with general laws all other facts of the evaluation world can be derived are important for the similarity relation. What remains to be worked out is how this idea can be made precise.

## 5.4 Premise semantics

### 5.4.1 A short history of premise semantics

That laws play a special role for the truth of *would have* conditionals is not at all a new idea. Particularly clear on this point are the proponents of *cotenability theories* for the meaning of conditionals. The cotenability approach dominated the thinking on the semantics of *would have* conditionals before the similarity approach came about. The founding work of this paradigm is Goodman (1965). He proposes that a *would have* conditional is true if the consequent can be derived from general laws and the antecedent plus a set of relevant conditions. Hence, the truth depend, in addition to the propositions expressed by the antecedent and the consequent, on two factors: (i) the set  $G$  of general laws (ii) a relevant set  $S$  of statements true of the evaluation world. According to Goodman, the problem of the meaning of *would have* conditionals is to specify these ingredients for a concrete example, in particular, to do so in a non-circular way.

In the seventies there was a strong opposition between the defenders of the similarity approach and the proponent of cotenability theories (see Fine 1975). This may seem surprising, because there is an obvious way to relate the two approaches. We can easily define a similarity relation based on cotenability the-

ory by proposing that the worlds most similar to the evaluation world are the worlds where all general laws hold plus the relevant conditions of the evaluation world. Thus, we can interpret cotenability theories as making the similarity relation precise. There is one branch of theories for the semantics of *would have* conditionals that follows this idea of combining the cotenability approach with the similarity theory. This is *premise semantics*, introduced by Veltman (1976) and Kratzer (1979, 1981a).<sup>9</sup> The basic idea behind premise semantics is this. We define a function, called by Veltman (1985) the *premise function*, that, given a set of possible worlds  $W$ , maps a member  $w$  of this set to a set  $P_w$  of propositions in  $W$ . Veltman (1985) interprets  $P_w$  as “your stock of beliefs in  $w$ ”; for Kratzer (1981) it is “everything which is the case in  $w$ ”. When a *would have* conditional is evaluated in  $w$  the interpreter tries to verify the consequent on those worlds where the antecedent and as many members of  $P_w$  as possible are true.<sup>10</sup>

The premise semantics rule for *would have* conditionals

Let  $W$  be a set of possible worlds and  $P$  be a premise function that maps a possible world  $w \in W$  on a set of propositions  $P_w$  in  $W$ . We say that a set of propositions  $S$  make a sentence  $\psi$  true, if every world  $w$  contained in all propositions in  $S$  makes  $\psi$  true. We say that a set of propositions  $S$  admits the sentence  $\psi$  if there is some world  $w$  contained in all elements of  $S$  that makes  $\psi$  true. A *would have* conditional is true in world  $w$  iff every maximal subset of  $P_w$  that admits the antecedent makes the consequent true.

It is easy to see that for every premise function you can find some order such that the premise semantics and the similarity approach evaluate exactly the same *would have* conditionals as true (see Veltman 1985 for discussion). The problem of specifying similarity now becomes a problem of specifying the premise function. Giving the initial idea of Kratzer that  $P_w$  is everything that is the case in  $w$ , one could suggest defining this function as the set of true propositions in  $w$ . However, this does not work. As Veltman shows, in this case the truth conditions of *would have* conditionals reduce to something very similar to strict implication (see Veltman, 1985, proposition II.65). Because the truth conditions of a strict conditional approach are not very satisfying, we have to dismiss this option. We can repair the approach based on Kratzer’s suggestion in two ways. First, we may restrict the facts about the actual world that are in  $P_w$ . Second, we may localize the problem in the way  $P_w$  is proposed to contribute to the meaning of *would*

---

<sup>9</sup>The name *premise semantics* has been introduced in Lewis (1981b).

<sup>10</sup>For more details see Veltman (1985). For simplicity I have chosen here a formulation of premise semantics that assumes that the premise function satisfies the limit assumption: each  $\psi$ -admitting subset of every set  $P_w$  is a subset of some maximal  $\psi$ -admitting subset of  $P_w$ . Without this assumption, the formulation of the last sentence in the definition would have to be: *A would have conditional is true in world  $w$  iff every subset of  $P_w$  that admits the antecedent can be extended to a set  $P'_w$  that makes the consequent true.*

*have* conditionals. Kratzer (1989) proposes that we have to take both options at once. Kratzer (1989) introduces some general restrictions on the set of facts of the actual world that may be relevant for the truth conditions of *would have* conditionals. But she also proposes some additional constraints on the subsets of  $P_w \cup A$  on which the truth of the consequent is checked, besides consistency. We will not go into the details of her analysis.<sup>11</sup> However, it is interesting to observe that she emphasizes that non-accidental generalizations, i.e. laws, are always in the set of propositions from which the consequent has to be derivable. Thus, she proposes that general laws are facts that cannot be given up by the similarity relation.

In reaction to the result mentioned above (Veltman 1985, proposition II.65) Veltman (1985) proposes that there have to be some asymmetries between the propositions selected by the premise function, i.e. the facts that count for similarity. Some may count more than others, some may not count at all. Hence, he suggests that the simple distinction in facts that count for similarity (those in  $P_w$ ) and facts that do not (those not in  $P_w$ ) that underlies the rule of premise semantics has to be given up. Now, we have to distinguish different classes of premises and describe their respective impact on similarity. Similar to Kratzer (1981a, 1989), one of these classes Veltman considers to be the class of laws we consider to be valid in a certain context. Elements of this class cannot be given up at any cost by similarity: “The role which laws – and other propositions we treat as such – play is important, since they determine which possible worlds can enter into the relation of comparative similarity and which cannot. Only those worlds in which the same laws hold as in the actual can.” (Veltman, 1985: 121). But Veltman (1985) realizes that more information about the actual world goes into the evaluation of *would have* conditionals besides what counts as law. Veltman (1985) tries to describe this additional information as those characteristics of the evaluation world that the interpreter is acquainted with. This would stand in direct tradition with the Ramsey recipe for the interpretation of conditionals. But he cites an example of Tichy (1976) that shows that this is not the correct way to. The following is a slight variation of the original example from Tichy.

*Consider a man - call him Jones - who is possessed of the following disposition as regards wearing his hat. If the man on the news predicts bad weather, Mr Jones invariably wears his hat the next day. A weather forecast in favor of fine weather, on the other hand, affects him neither way: in this case he puts his hat on or leaves it on the peg, completely at random. Suppose, moreover, that yesterday bad weather was prognosed, so Jones is wearing his hat. In this case, ...*

- (66) If the weather forecast had been in favor of fine weather, Jones would have been wearing his hat.

---

<sup>11</sup>Notice, that there exist some strong objections against her proposal (see Kanazawa et al. 2005).

In this context *Jones wears his hat* is a fact of the actual world that we are aware of. There is no reason why when making minimal amendments to what we are aware of concerning the world described in this context in order to make the antecedent of (66) true, this fact should be given up. But then the *would have* conditional (66) comes out as true, while intuitively it is false.

Hence, for some reason the fact *Jones wears his hat* seems to be excluded from the facts that count for similarity. Thus, the criterion of awareness does not work. In his book from 1985 Veltman closes with admitting that he does not know how to distinguish between facts that do count and those that do not.

Let me add a final remark on this example. One may again be tempted to propose that temporal properties of the involved facts are relevant for similarity in this case. One may propose, for instance, that facts about the future of the evaluation time of the antecedent in general do not count for similarity. Then that Mr. Jones is wearing his hat in the evaluation world would have no impact anymore and one would correctly predict that (66) is false in the given context. Example (67) shows that this will not do. The *would have* conditional (67) is intuitively true. But that means that the outcome of the chance event, that lies in the future of my betting, has to count for similarity. This example clearly shows that the future of the evaluation time of the antecedent matters for similarity.

*A coin is going to be thrown and you have bet \$5 on heads. Fortunately, heads comes up and you win. You say*

(67) If I had bet on tails, I would have lost.

### 5.4.2 Explaining Mr. Jones with premise semantics

Veltman (2005) comes forward with a new approach to the meaning of *would have* conditionals that solves the Tichy puzzle. This approach is again an instantiation of premise semantics. As proposed above, Veltman (2005) distinguishes two sets of premises. First, there is a contextually given set of laws taken to be valid in the evaluation world. The most similar antecedent worlds have to fulfill all of these laws. But there is also a second set of premises, which are singular facts about the evaluation world  $w$ . Veltman (2005) calls this set of facts the *basis* of  $w$ . The basis is now not described as the facts the interpreter is acquainted with, but as a minimal set of propositions that, together with the laws, completely determine the evaluation world  $w$ . In other words, from the basis and the laws every other facts of  $w$  has to be derivable.<sup>12</sup> The basis goes into the calculation of similarity in the standard way: a world  $w'$  counts more similar to  $w$  the more elements of some basis of  $w$  it makes true.

---

<sup>12</sup>There can be more than one basis for a world.

Let us sketch this approach a bit more precisely.<sup>13</sup> Let  $U$  be the worlds consistent with the general laws taken to hold in the context of evaluation. We write  $w \models A$ , if  $A$  is true in  $w$ . For a set of sentences  $S$  we write  $w \models S$  if every element of  $S$  is true in  $w$ . Then we can describe Veltman's (2005) proposal for the truth conditions of *would have* conditionals as follows.

Veltman's (2005) interpretation rule for *would have* conditionals  
 A *would have* conditional with antecedent  $A$  and consequent  $C$  is true in  $w$  iff  $C$  holds in those worlds  $w'$  that fulfill the following three conditions.

- $w'$  makes the antecedent true:  $w' \models A$ ,
- $w'$  obeys all laws:  $w' \in U$ , and
- for some basis  $b$  of  $w$  there is a maximal subset  $s \subseteq b$  such that for some world  $w'' \in U : w'' \models s \cup \{A\}$  and  $w' \models s$ .

Ignoring the possibility of multiple bases, we can say that Veltman's (2005) approach splits the premise function from the last section into two subfunctions: one function selects the set of laws and the other the basis of the evaluation world. The facts in the two sets are proposed to be of different importance to similarity. Laws have to be observed always; they can never be given up. The interpreter tries to keep as many facts of a basis as is consistent with the antecedent and the laws. Notice the similarity of this approach to what was proposed by Goodman. The truth of a *would have* conditional is proposed to be determined by the contextually relevant laws and a set of relevant singular facts of the evaluation world. The central contributions of Veltman (2005) are (i) to combine central ideas of the cotenability theory and the similarity approach, and (ii) to give a precise description of the second set of premises : the relevant singular facts of the evaluation world.

To illustrate the power of the approach let us finally show how it accounts for the Tichy example (66) discussed in section 5.4.1. Let *bad* stand for the sentence that the weather forecast has predicted bad weather and *hat* for the sentence that Mr. Jones is wearing his hat. The evaluation world in which we want to consider the truth of the conditional (66) is such that the weather has been predicted to be bad and Mr. Jones is wearing his hat. In the context of interpretation for the example a general law is introduced:  $bad \rightarrow hat$ , i.e. *always if the weather forecast is in favor of bad weather, Mr. Jones wears his hat*. To calculate the truth of the conditional we have to determine the bases of the evaluation world. There is only one basis, the set  $\{bad\}$ . The fact *hat* does not count as an element of the basis, because it can be derived from the fact *bad* and the general law  $bad \rightarrow hat$ . What are now the worlds  $w'$  on which the consequent has to be true? First, these

---

<sup>13</sup>For the formal details the interested reader is referred to the original paper.

worlds have to make the antecedent true:  $w' \models \neg bad$ . Second, they have to obey the laws:  $w' \models bad \rightarrow hat$ . Finally,  $w'$  has to make a maximal subset of the basis  $\{bad\}$  true. Because  $w' \models \neg bad$  this maximal subset is the empty set. From this we can conclude that there are two type of worlds fulfilling all these conditions: worlds where the weather is predicted to be fine and Mr. Jones wears his hat, but also worlds where the weather is predicted to be fine and Mr. Jones has left his hat home. Because of this second class of worlds, Veltman (2005) predicts that (66) is false, as intended.

### 5.4.3 Problems of the approach

Veltman (2005) proposes a surprisingly simple and intuitively appealing treatment of the meaning of *would have* conditionals. However, there are still some open problems for this account. One question, that is already discussed in Veltman (2005), is the issue of epistemic readings of *would have* conditionals. Veltman uses the following example to illustrate the epistemic reading (Veltman 2005:174).

“The duchess has been murdered, and you are supposed to find the murderer. At some point only the butler and the gardener are left as suspects. At this point you believe

(68) If the butler did not kill her, the gardener did.

Still, somewhat later – after you found out convincing evidence showing that the butler did it, and that the gardener had nothing to do with it – you get in a state, in which you will *reject* the sentence

(69) If the butler had not killed her, the gardener would have.”

Veltman’s (2005) approach predicts for (69) the intuition reported in the example, namely that the *would have* conditional is false. However, many people do not agree with this judgment, as similar examples in the literature show (see, for instance, the extensive discussion on the Hamburger example, introduced in Hansson 1989). Even Veltman himself gives an example in his dissertation where he judges a conditional very similar to (69) as true (see Veltman 1985: 217). Hence, there is substantial evidence that there exists a reading of (69) that takes the conditional to be true in the described context. The approach of Veltman (2005) as it stands cannot deal with this reading.

Veltman (2005) also discusses another type of example his approach has troubles with. These are *would have* conditionals that are based on a law that concludes from the truth of two premises to the truth of the consequent:  $prem1 \wedge prem2 \rightarrow cons$ . The critical predictions turn up when in the evaluation world the first premise is true, the second false, and the consequent false as well. In such a context a *would have* conditional *If the second premise had been true as*

*well, the consequent would have been true* is sometimes intuitively true. Veltman's approach, however, in general predicts that in such a situation the conditional is false. The reason is that the basis of the described evaluation world consists of the true premise and the false consequent. His evaluation conditions for conditionals predict that the closest worlds where the antecedent is true are worlds where the first premise is true, and hence, the consequent true, but also worlds where the consequent is false and consequently the first premise false as well. On the first type of world the consequent of the conditional *If the second premise had been true as well, the consequent would have been true* is true, but on the second type it is false. Consequently, the conditional is predicted to be false. This problem can, for instance, be illustrated with the following example from Lifschitz. Intuitively, the conditional (70) is true in the described context. However, Veltman (2005) predicts it to be false for the reasons just discussed.

*Suppose there is a circuit such that the light is on exactly when both switches are in the same position (up or not up). At the moment switch one is down, switch two is up and the lamp is out. Now consider the following would have conditional:*

(70) If switch one had been up, the lamp would have been on.

There are other examples the approach has difficulties to treat. They again suggest that time, in particular the past, plays a role for similarity. Let us consider, for instance, a variation of the famous Kennedy example from Adams (1975). Adams used the following two sentences to illustrate that there is a truth-conditional difference between indicative and subjunctive conditionals. Most people agree that (71a) is true while they reject at the same time (71b).

- (71) a. If Oswald didn't killed Kennedy, somebody else did.  
       b. If Oswald hadn't killed Kennedy, somebody else would have.

Now, let us reconsider the *would have* conditional (71b) in the following context.

The Kennedy-conspiracy example

*Assume that there was a big conspiracy to kill Kennedy. The participants planned the assassination attempt of Oswald, but also a whole sequence of other attempts carried out by different people. Just by accident Oswald was the first one to succeed in killing Kennedy.*

In this context (71b) is or can be imagined to be true. The approach of Veltman (2005), however, predicts this *would have* conditional to be false. The reason is that the fact that there is a conspiracy in the evaluation world and the



fact that Oswald killed Kennedy and nobody else did are taken by this approach to be of equal importance for the similarity relation. Together with the fact that Oswald did kill Kennedy both facts mentioned above define a basis for the evaluation world. But that means that among the antecedent worlds closest to the evaluation world there are worlds where somebody else killed Kennedy as well as worlds where no conspiracy took place. The later type of worlds falsifies the consequent of (71b).

If we want to account for the Kennedy-conspiracy example within the similarity framework, we somehow have to find a difference between the fact that there is a conspiracy and the fact that it was Oswald who murdered Kennedy. A proposal that immediately suggests itself is to make temporal properties of the two facts responsible for their different impact on similarity. Given the context of (71b), the conspiracy is set up before Oswald kills Kennedy. Hence, one might again suggest that facts that are further in the past count more for similarity. But we have seen in section 5.3.2 that it is not at all clear how this can be made precise without leading to new problems. In the next section we will discuss a totally different way to approach the meaning of *would have* conditionals that is able to account for this type of counterexample to the approach of Veltman (2005) without direct reference to time.

## 5.5 Counterfactuals in causal networks

### 5.5.1 The general ideas

The next theory for the meaning of *would have* conditionals we are going to discuss has been proposed in a book on causation. For linguists it may be surprising to find a semantic theory in such a place. But philosophers familiar with the subject will of course recall the central role counterfactuals play in the philosophical discourse on causation. Think, for instance, of Lewis' claim that causality can and should be explained in terms of counterfactuals. Even for philosophers it might be surprising that this book is written by a computer scientist. However, there is a growing awareness of the central role causation plays for empirical, particularly statistical models, which has spurred a lot of activity in developing formal and mathematical models for causation in these areas.

The author of the book, Judea Pearl, agrees with the tenor of cotenability theory, premise semantics and another tradition of theories for the meaning of *would have* conditionals: the probabilistic approach (see Adams 1975, 1976 and Skyrms 1980, 1981, 1984, 1994): if we want to formalize the meaning of *would have* conditionals<sup>14</sup>, we need to make a distinction between relationships in the world that are stable, invariant under certain changes – we might call them laws

---

<sup>14</sup>Pearl (2000) speaks of counterfactuals and not of *would have* conditionals. However, I will apply his theory to my notion of *would have* conditionals.

– and accidental information over the world. Both are relevant for the interpretation of these sentences, but in different ways. According to Pearl (2000) many earlier attempts to formalize the meaning of *would have* conditionals fail because the formal language they use is not appropriate to make this distinction. This is, following him, in particular a problem for the classical logical approaches to counterfactuals (see Pearl 2000: 224), but also, even though to a less degree, for approaches using probability theory.<sup>15</sup> Besides the inability to make a natural distinction between facts and laws, Pearl (2000) sees another systematic problem with applying classical logic or probability theory to a description of the meaning of counterfactuals. Counterfactuals “... deal with changes that occur in the external world rather than with changes in our beliefs about a statical world.” (Pearl 2000: 203). According to him classical logic and probability theory are designed to account only for the latter. Therefore we need new structures to deal with counterfactuals: *causal models*. “Causal models encode and distinguish information about external changes through an explicit representation of the mechanisms that are altered in such changes.” (Pearl 2000: 203).

Pearl (2000) proposes that counterfactuals are evaluated by thinking about the consequences it has if you actively encroach upon reality and manipulate the value of certain variables. More particularly, he proposes that to make the antecedent true we have to cut off the causal history leading to the falsity of the antecedent and simply stipulate its truth (without caring about how this truth came about). This manipulative, or interventional, thinking about counterfactual reasoning fits nicely with an old idea in the literature on the semantics of *would have* conditionals. “Often, indeed, we seem to reason in a way that takes it for granted that the past is counterfactually independent of the present: that is, that even if the present were different, the past would be just as it actually is.” (Lewis 1979: 455-6). It is not at all obvious, how this idea can be formalized. The answer Pearl (2000) provides to this question will now be studied in some detail.

### 5.5.2 The formalization

The presentation of Pearl’s (2000) theory given below deviates from the way Pearl introduces his ideas. The main motivation for choosing a different description of these ideas is to make the theory much more transparent for readers with a background in logic or formal semantics. However, the changes that are made only affect the formulation of the approach, not its substance.

Central to the whole approach stands the notion of a causal model. A causal model describes a fragment of the causal dependencies that govern reality. It

---

<sup>15</sup>In probability theory there is, according to Pearl (2000), a natural distinction between laws and facts: “Facts express ordinary propositions and hence can obtain probabilities and can be conditioned on [ ... ]; laws, on the other hand, are expressed as conditional probability sentences and hence should not be assigned probabilities and cannot be conditioned on.” (Pearl 2000: 224).

consists, first, of a set of variables. Pearl allows any kinds of variables: variables for degrees on a thermometer, variables for the color of your hair, etc. For our purposes it is sufficient to consider in place of variables a set of proposition letters  $\mathcal{P}$ . The set  $\mathcal{P}$  is split into two subclasses: a set  $B$  of background variables, which are taken to be determined by (causal) processes external to the causal model under discussion, and the endogenous variables  $E = \mathcal{P} - B$ , that depend causally on the value of other variables of the model, which may either be background variables or other endogenous variables. The exact character of the dependence is described in the second ingredient of a causal model. This is a function  $F$  that associates every endogenous variable with a formula in propositional logic that may involve all other variables. Given an endogenous variable  $Y$  the corresponding formula  $F(Y)$  determines the value of  $Y$  dependent on the value of the variable occurring in the formula.

**5.5.1. DEFINITION.** (Causal models according to Pearl)

Let  $\mathcal{P}$  be a finite set of proposition letters and  $\mathcal{L}$  the language you obtain when closing  $\mathcal{P}$  under negation and conjunction. A *causal model* for  $\mathcal{P}$  is a triple  $M = \langle B, E, F \rangle$ , where

- i.  $B \subseteq \mathcal{P}$  are called *background* variables;
- ii.  $E = \mathcal{P} - B$  are called *endogenous* variables; and
- iii.  $F$  is a function  $F : (\mathcal{P} - B) \longrightarrow \mathcal{L}$  that is rooted in  $B$ .

We want the function  $F$  to allow us to determine the value of all endogenous variables based on an interpretation of the background variables. This does not mean that in all formulas associated with endogenous variables only background variables may occur. What we want is that if you go backward and check which variables are used in the formula and which variables are used in the formulas associated with these variables etc. at some point all these lines will lead to background variables. This will be warranted by the condition that  $F$  is *rooted* in  $B$ .

**5.5.2. DEFINITION.** (Rootedness)

Let  $\mathcal{P}$  be a finite set of proposition letters and  $\mathcal{L}$  the language you obtain when closing  $\mathcal{P}$  under negation and conjunction. Let  $M = \langle B, E, F \rangle$  be a causal model. We introduce a binary relation  $R_F$  on the set of proposition letters  $\mathcal{P}$ .  $R_F(X, Y)$  holds, if  $X$  occurs in  $F(Y)$ .<sup>16</sup> Let  $R_F^T$  be the transitive closure of  $R_F$ . The  $R_F$ -minima of a letter  $Y \in \mathcal{P}$ ,  $Min_{R_F}(Y)$ , are defined as follows:

$$MIN_{R_F}(Y) = \{X \in \mathcal{P} \mid R_F^T(X, Y) \ \& \ \neg \exists Z \in \mathcal{P} : R_F^T(Z, X)\}$$

We say that  $F$  is *rooted* in  $B$  if  $R_F^T$  is acyclic and  $\forall Y \in \mathcal{P} - B : Min_{R_F}(Y) \subseteq B$ .

---

<sup>16</sup>By the definition of  $F$  this means that  $Y \in \mathcal{P} - B$ .

The notion of rootedness consists of two conditions. First, we demand that  $R_F^T$  is acyclic. This comes down to the claim that the effect of some cause cannot be causally responsible for the cause, even in an indirect way. The second condition,  $\forall X \in \mathcal{P} - B : \text{Min}_{R_F}(X) \subseteq B$ , demands that everything starts with the background variables. In other words, if you move backward along the relation  $R_F$  you will always end up at some element in  $B$ .

To illustrate the working of these definitions we will discuss an example. Remember Lifschitz' circuit example from section 5.4.3.

*Suppose there is a circuit such that the light is on exactly when both switches are in the same position (up or not up). At the moment switch one is down, switch two is up and the lamp is out.*

We want to give a causal model that describes the causal dependencies of the given contexts. The most straightforward way to go is to distinguish two background variables,  $S_1$   $S_2$ , and one endogenous variable  $L$ . All three are proposition letters, taking as value the truth values 1 or 0.  $S_1$  is set to 1 if switch one is up and to 0 otherwise. Analogously,  $S_2$  is connected to the position of switch two. The variable  $L$  is set to 1 if the lamp is on and to 0 if it is off. Finally, the function  $F$  maps  $L$  to some formula of the language  $\mathcal{L}$  generated by the set of proposition letters  $\mathcal{P} = \{S_1, S_2, L\}$ , that is to express the causal dependency of the state of the lamp from the position of the switches: the lamp is on if and only if the switches are in the same position. This should be, of course the formula  $S_1 \leftrightarrow S_2$ .<sup>17</sup> In sum, we can model the causal dependencies of the Lifschitz' example with the causal model  $M = \langle B, E, F \rangle$ , where  $B = \{S_1, S_2\}$ ,  $E = \{L\}$ , and  $F(L) = S_1 \leftrightarrow S_2$ .

Every causal model  $M$  can be associated with a directed acyclic graph,  $G(M)$ . This representation can be very useful to get an intuitive understanding of a causal model. The graph is defined by letting each node correspond to a proposition letter and introduce a directed edge from letters  $X$  to  $Y$  if  $R_F(X, Y)$  holds. Keep in mind, however, that a graph merely identifies the variables that have direct influence on each endogenous variable; the graph does not specify the exact nature of the dependency. The graph for the model of the Lifschitz example we have just given is shown in figure 5.1.

An important consequence of the definition of a causal model is that it imposes a strong condition on causal dependencies: if you know the causes of a certain effect and you know the values these causes have, then you can always determine the value of the effect. It is not possible that for some valuation of the causes the value of the effect is not determined. Formally, this is realized by identifying the value of some endogenous variable  $Y$  (the effect) with the truth value of the formula associated with  $Y$  by  $F(Y)$ . As long as the propositional letters occurring in this formula (the causes) have a truth value, the formula will have a

<sup>17</sup>Where  $A \leftrightarrow B$  abbreviates  $\neg(A \wedge \neg B) \wedge \neg(B \wedge \neg A)$ .

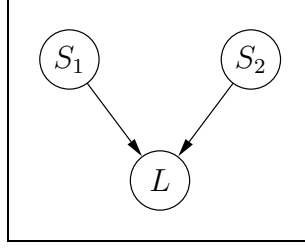


Figure 5.1: A graph for the Lifschitz Example

truth value – and, in consequence, also  $Y$ . Pearl calls this property of his causal models *determinism*. Because of this property and the property of rootedness, a causal model allows you to determine the value of every endogenous variable given some valuation of the background variables. Hence, given a causal model  $M$  we can extend every interpretation  $I$  of the letters in  $B$  first to an interpretation of all proposition letters  $\mathcal{P}$  and then to an interpretation of the language  $\mathcal{L}$  generated from  $\mathcal{P}$ . This is formally specified in the next definition.<sup>18</sup>

### 5.5.3. DEFINITION. (Truth values for $\mathcal{L}$ generated by a causal model)

Let  $\mathcal{P}$  be a set of proposition letters and  $\mathcal{L}$  the closure of  $\mathcal{P}$  under conjunction and negation. Furthermore, let  $M = \langle B, E, F \rangle$  be a causal model for  $\mathcal{P}$  and  $I : B \rightarrow \{0, 1\}$  an interpretation of the background variables of  $M$ . For arbitrary  $\psi \in \mathcal{L}$  we define the interpretation of  $\psi$  with respect to  $M$  and  $I$ ,  $\llbracket \psi \rrbracket^{M,I}$  recursively as follows.

- $\llbracket \psi \rrbracket^{M,I} = I(\psi)$ , if  $\psi \in B$ ,
- $\llbracket \psi \rrbracket^{M,I} = \llbracket F(\psi) \rrbracket^{M,I}$ , if  $\psi \in \mathcal{P} - B$ ,
- $\llbracket \neg\psi \rrbracket^{M,I} = 1$ , iff  $\llbracket \psi \rrbracket^{M,I} = 0$ , and
- $\llbracket \psi \wedge \phi \rrbracket^{M,I} = 1$ , iff  $\llbracket \psi \rrbracket^{M,I} = 1$  and  $\llbracket \phi \rrbracket^{M,I} = 1$ .

In addition to elements of  $\mathcal{L}$ , we also want truth conditions for *would have* conditionals. To express the conditional within this formal framework we use the connective  $\succ$ . Thus, a *would have* conditional of  $\mathcal{L}$  is formally a sentence  $\psi \succ \phi$  with  $\psi, \phi \in \mathcal{L}$ . An important limitation of the approach of Pearl is, as we will see, that the antecedent of such conditionals is restricted to conjunctions of literals of

<sup>18</sup>As a side-mark: one might wonder whether, given this dependency of the endogenous variables on the background variables, we can reduce the complexity of a causal model and let  $F$  assign to every endogenous variable  $Y$  a formula containing only background variables – those background variables in  $\text{Min}_{R_F}(X)$ . It might seem as if there is nothing we can lose this way. But in fact we would lose important information. Such a notion of a causal model would not tell us which of the endogenous variables causally depend on others. But as we will see, this information is crucial to model the truth conditions of *would have* conditionals.

$\mathcal{L}$ , i.e. elements of  $\mathcal{P}$  or the negation thereof. Furthermore, the approach of Pearl works only for antecedents that are made up entirely of endogenous variables. This means that not all sentences  $\psi \succ \phi$  are interpretable according to Pearl (2000). We will see that this leads to problems when we come to the discussion of concrete examples. The basic interpretation rule of Pearl (2000) for *would have* conditionals follows the general ideas described in the introduction. A sentence  $\psi \succ \phi$  is said to be true with respect to a causal model  $M$  and an interpretation function  $I$ , if the consequent  $\phi$  is true with respect to the same interpretation function  $I$  and the causal model  $M_\psi$  you obtain by manipulating  $M$  to force the truth of the antecedent  $\psi$  *by law*.

**5.5.4. DEFINITION.** (Pearl's truth conditions for *would have* conditionals)

Let  $\mathcal{P}$  be a set of proposition letters and  $\mathcal{L}$  the closure of  $\mathcal{P}$  under conjunction and negation. Let  $M = \langle B, E, F \rangle$  be a causal model for  $\mathcal{P}$  and  $I$  be a function from  $B$  to  $\{0, 1\}$ . For  $\psi, \phi \in \mathcal{L}$ , where  $\psi$  is a conjunction of literals of elements of  $E = \mathcal{P} - B$ , we define

$$\llbracket \psi \succ \phi \rrbracket^{M, I} = 1, \text{ iff } \llbracket \phi \rrbracket^{M_\psi, I} = 1.$$

The question that still has to be answered is how to define  $M_\psi$ . Pearl proposes that in  $M_\psi = \langle B, E, F' \rangle$   $F'$  associates the variables occurring in  $\psi$  with a different formula than does the function  $F$  of the original model.<sup>19</sup> For every endogenous variable  $P$  occurring in  $\psi$ ,  $F'$  maps  $P$  to  $\top$  if  $P$  is among the positive literals of  $\psi$ <sup>20</sup>, and  $F'$  maps  $P$  to  $\perp$  if  $P$  is among the negative literals of  $\psi$ <sup>21</sup>. At this point it becomes crucial that  $\psi$ , the antecedent of a Pearl conditional, is a conjunction of literals.

**5.5.5. DEFINITION.** (Intervention by Pearl)

Let  $\mathcal{P}$  be a set of proposition letters and  $\mathcal{L}$  the closure of  $\mathcal{P}$  under conjunction and negation. Let  $M = \langle B, E, F \rangle$  be a causal model for  $\mathcal{L}$ . For  $\psi \in \mathcal{L}$  where  $\psi$  is a conjunction of literals of elements of  $E = \mathcal{P} - B$ , the model  $M_\psi = \langle B', E', F' \rangle$  is defined as follows:

- i.  $B' = B$ ,
- ii.  $E' = E$ ,
- iii.  $\forall X \in E$ : if  $X$  does not occur in  $\psi$ , then  $F'(X) = F(X)$ ,
- iv.  $\forall X \in E$ : if  $X$  occurs positively in  $\psi$ , then  $F'(X) = \top$ , and

---

<sup>19</sup>Remember that  $\psi$  is made up entirely of endogenous variables.

<sup>20</sup>That means that  $P$  occurs in  $\psi$  but not  $\neg P$ . We also say in this case that  $P$  *occurs positively* in  $\psi$ .

<sup>21</sup>That means that  $\neg P$  occurs as literal in  $\psi$ . We also say in this case that  $P$  *occurs negatively* in  $\psi$ .

v.  $\forall X \in E$ : if  $X$  occurs negatively in  $\psi$ , then  $F'(X) = \perp$ .

Let us again illustrate these definitions with the Lifschitz example. We want to evaluate the *would have* conditional (72) in the context repeated below.

*Suppose there is a circuit such that the light is on exactly when both switches are in the same position (up or not up). At the moment switch one is down, switch two is up and the lamp is out. Now consider the following would have conditional:*

(72) If switch one had been up, the lamp would have been on.

The first thing that catches the eyes is that we cannot use the model given in figure 5.1 to evaluate the conditional. If we used this model, we would have to check whether  $\llbracket S_1 \succ L \rrbracket^{M,I} = 1$ , where  $I$  maps  $S_1$  to 0 and  $S_2$  to 1. But the variable  $S_1$  is a background variable in  $M$ . As said above, Pearl's truth conditions do not work for background variables in the antecedent. We have to turn  $S_1$  into an endogenous variable without changing the causal functionality of the model. The easiest way to go is to move  $S_1$  to the endogenous variables, introduce a new background variable  $U$ , and extend  $F$  to  $S_1$  with the condition  $F(S_1) = U$ . Hence, the model becomes  $M = \langle B, E, F \rangle$ , where  $B = \{U, S_2\}$ ,  $E = \{S_1, L\}$ , and  $F(L) = S_1 \leftrightarrow S_2$ ,  $F(S_1) = U$ . The graph of this model is given in figure 5.2.

The introduction of the additional variable  $U$  does not change anything with respect to the already existing causal dependencies of the system. Pearl's approach needs  $U$  to be able to manipulate the value of  $S_1$  via causal laws. This is not possible if  $S_1$  is a background variable, because then its value is determined by the interpretation function  $I$ . This said, it should become clear that we could also introduce such a dummy variable for  $S_2$  without any effects on the predictions made. We refrain from this step, because there is no need for such an unmotivated addition to the model.

With this model at hand we can now evaluate whether the conditional (72) holds in the given context. The interpretation function  $I$  for the background variables of  $M$  that fits this example assigns 1 to  $S_2$  and 0 to  $U$ . We want to calculate  $\llbracket S_1 \succ L \rrbracket^{M,I}$ , which is, according to definition 5.5.4, 1 if and only if  $\llbracket L \rrbracket^{M_{S_1},I} = 1$ . The central step is to calculate the model  $M_{S_1}$  with the help of definition 5.5.5. We obtain  $M_{S_1} = \langle B', E', F' \rangle$  with  $B' = B$ ,  $E' = E$ ,  $F'(L) = F(L)$ , and  $F'(S_1) = \top$ . The truth of  $\llbracket L \rrbracket^{M_{S_1},I}$  can be established by checking the truth conditions for  $\mathcal{L}$  laid down in definition 5.5.3 (see the calculation given below). (72) comes out as true according to Pearl's theory, as intended.

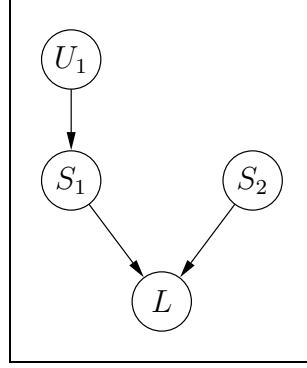


Figure 5.2: The extended graph for the Lifschitz Example

$$\begin{aligned}
\llbracket L \rrbracket^{M_{S_1}, I} = 1 & \text{ iff } \llbracket F'(L) \rrbracket^{M_{S_1}, I} = 1 \\
& \text{ iff } \llbracket S_1 \leftrightarrow S_2 \rrbracket^{M_{S_1}, I} = 1 \\
& \text{ iff } \llbracket S_1 \rrbracket^{M_{S_1}, I} = 1 \Leftrightarrow \llbracket S_2 \rrbracket^{M_{S_1}, I} = 1 \\
& \text{ iff } \llbracket F'(S_1) \rrbracket^{M_{S_1}, I} = 1 \Leftrightarrow I(S_2) = 1 \\
& \text{ iff } \llbracket \top \rrbracket^{M_{S_1}, I} = 1 \Leftrightarrow I(S_2) = 1 \\
& \text{ iff } I(S_2) = 1
\end{aligned}$$

### 5.5.3 More examples

To further illustrate the working and the power of this approach let us discuss two more examples. We will start with the Kennedy conspiracy example that has been introduced in section 5.4.3 as problematic for the approach of Veltman (2005).

*Assume that there was a big conspiracy to kill Kennedy. The participants planned the assassination attempt of Oswald, but also a whole sequence of other attempts carried out by different people. Just by accident Oswald was the first one to succeed in killing Kennedy.*

(73) If Oswald hadn't killed Kennedy, someone else would have.

In the following I will make the simplifying assumption that the attempt of Oswald to kill Kennedy was actually the assassination attempt that was scheduled first. This simplification is not essential to the treatment of the example. The natural set of proposition letters that we have to introduce to capture this example is  $\mathcal{P} = \{K_1, K_2, D\}$ , where  $K_1$  represents that Oswald kills Kennedy in his assassination attempt,  $K_2$  represents that some assassination attempt scheduled later is successful, hence, someone else kills Kennedy, and  $D$  represents that



Kennedy dies. As in the last example we will need a dummy letter  $U$  to turn  $K_1$  into an endogenous variable. Otherwise, Pearl's approach is not able to evaluate the conditional  $\neg K_1 \succ K_2$ . The critical part is to formalize the conspiracy idea within a causal network. We propose that the context can be described using two causal dependencies in addition to the dummy dependency  $F(K_1) = U$ . First, if some assassination attempt succeeds, then Kennedy dies:  $F(D) = K_1 \vee K_2$ . Second, a later assassination will take place and will only take place, if Oswald's attempt to kill Kennedy fails:  $F(K_2) = \neg K_1$ . The complete causal model with graph is given in figure 5.3.

$$\begin{aligned}
 M &= \langle B, E, F \rangle \\
 B &= \{U\} \\
 E &= \{K_1, K_2, D\} \\
 F(K_1) &= U \\
 F(K_2) &= \neg K_1 \\
 F(D) &= K_1 \vee K_2
 \end{aligned}$$

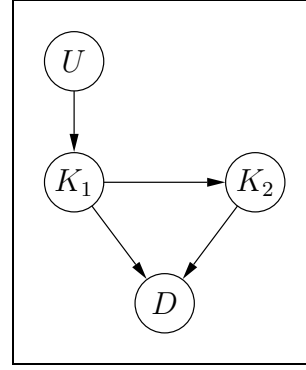


Figure 5.3: A causal model for the Kennedy Example

The interpretation function  $I$  of the background variable  $U$  that fits the context of the Kennedy conspiracy example is  $I(U) = 1$ : in the evaluation world Oswald in fact killed Kennedy; hence,  $K_1$  is true and, thus,  $U$  has to be true. Let us now evaluate the conditional (73) given this model and interpretation function. To check whether  $\llbracket \neg K_1 \succ K_2 \rrbracket^{M,I} = 1$ , we have to calculate  $M_{\neg K_1}$ . We obtain  $M_{\neg K_1} = \langle B', E', F' \rangle$ , with  $B' = B$ ,  $E' = E$ ,  $F'(D) = F(D)$ ,  $F'(K_2) = F(K_2)$ , and  $F'(K_1) = \perp$ . The following calculation shows that the conditional (73) comes out as true.

$$\begin{aligned}
 \llbracket K_2 \rrbracket^{M_{\neg K_1}, I} = 1 & \text{ iff } \llbracket F'(K_2) \rrbracket^{M_{\neg K_1}, I} = 1 \\
 & \text{ iff } \llbracket \neg K_1 \rrbracket^{M_{\neg K_1}, I} = 1 \\
 & \text{ iff } \llbracket K_1 \rrbracket^{M_{\neg K_1}, I} = 0 \\
 & \text{ iff } \llbracket F'(K_1) \rrbracket^{M_{\neg K_1}, I} = 0 \\
 & \text{ iff } \llbracket \perp \rrbracket^{M_{\neg K_1}, I} = 0
 \end{aligned}$$

As a second example we will discuss the famous shooting squad example from the literature on causality.

*There is a court, an officer, two riflemen and a prisoner. If the court orders the execution, then the officer will give a signal to the riflemen. If the officer gives the signal to the riflemen, then the riflemen will*

*shoot. If a rifleman shoots, then the prisoner will die. The court orders the execution. The officer gives the signal. The riflemen both shoot. The prisoner dies.*

(74) (Even) if rifleman A hadn't shot, the prisoner would have died.

In this case we have to distinguish the following proposition letters:  $C$  for the court orders the execution,  $O$  for the officer gives the signal,  $R_1$  for rifleman one shoots,  $R_2$  for rifleman two shoots, and  $P$  for the prisoner dies. The causal model described by this context is given in figure 5.4. The interpretation of the background variable  $C$  that fits the given context is  $I(C) = 1$ . Now, we want to calculate whether  $\llbracket \neg R_1 \succ P \rrbracket^{M,I} = 1$ . This involves, first, the calculation of the model  $M_{\neg R_1} = \langle B', E', F' \rangle$ . Definition 5.5.5 tells us that  $B' = B$ ,  $E' = E$ ,  $F'(O) = F(O)$ ,  $F'(R_2) = F(R_2)$ ,  $F'(P) = F(P)$ , and  $F'(R_1) = \perp$ . The calculation below shows that  $\llbracket P \rrbracket^{M_{\neg R_1}, I} = 1$  and, hence, that the conditional comes out as true, as intended.

$$\begin{aligned}
 \mathcal{P} &= \{C, O, R_1, R_2, P\}, \\
 M &= \langle B, E, F \rangle, \\
 B &= \{C\} \\
 E &= \{O, R_1, R_2, P\}, \\
 F(O) &= C, \\
 F(R_1) &= O, \\
 F(R_2) &= O, \\
 F(P) &= R_1 \vee R_2.
 \end{aligned}$$

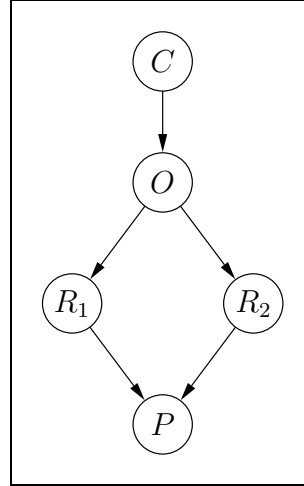


Figure 5.4: A causal model for the shooting squad example

$$\begin{aligned}
 \llbracket P \rrbracket^{M_{\neg R_1}, I} = 1 & \text{ iff } \llbracket F'(P) \rrbracket^{M_{\neg R_1}, I} = 1 \\
 & \text{ iff } \llbracket R_1 \vee R_2 \rrbracket^{M_{\neg R_1}, I} = 1 \\
 & \text{ iff } \llbracket R_1 \rrbracket^{M_{\neg R_1}, I} = 1 \text{ or } \llbracket R_2 \rrbracket^{M_{\neg R_1}, I} = 1 \\
 & \text{ iff } \llbracket F'(R_1) \rrbracket^{M_{\neg R_1}, I} = 1 \text{ or } \llbracket F'(R_2) \rrbracket^{M_{\neg R_1}, I} = 1 \\
 & \text{ iff } \llbracket \perp \rrbracket^{M_{\neg R_1}, I} = 1 \text{ or } \llbracket O \rrbracket^{M_{\neg R_1}, I} = 1 \\
 & \text{ iff } \llbracket O \rrbracket^{M_{\neg R_1}, I} = 1 \\
 & \text{ iff } \llbracket F'(O) \rrbracket^{M_{\neg R_1}, I} = 1 \\
 & \text{ iff } \llbracket C \rrbracket^{M_{\neg R_1}, I} = 1 \\
 & \text{ iff } I(C) = 1
 \end{aligned}$$

### 5.5.4 Discussion

Pearl can account for many examples that are problematic for other theories of *would have* conditionals. In general, the proposal accounts correctly for all non-backtracking examples where a directed causal path from antecedent to consequent can be assumed. In particular, it correctly describes the intuitions for the circuit example and the Kennedy example that have been found problematic for Veltman (2005) (see section 5.4.3), the Nixon example we discussed in connection with the future similarity objection (see section 5.3.2)<sup>22</sup> and the coin example mentioned in section 5.4.1.

Pearl's approach did not come out of the blue. The idea to describe the meaning of counterfactuals based on causal dependencies and also to describe the latter as functional dependencies between variables has been around for quite some time (see, for instance, Simon & Rescher 1966). On the first view, this approach to counterfactuals differs clearly from the traditional philosophical or linguistic approaches for counterfactuals or conditionals in general and the similarity approach of Stalnaker (1969) and Lewis (1973) we have discussed in this chapter. Nevertheless there is an intuitive connection between Pearl's theory and the similarity approach. This becomes particular obvious if one looks at Pearl's informal description of the interpretation strategy of counterfactuals he proposes. He describes the construction of the antecedent models  $M_\psi$  as follows: "... [Intervention, the author] bends the course of history (minimally) to comply with the hypothetical condition [of the truth of the antecedent, the author]" (Pearl, 2000: 37). This suggests that one can read Pearl's interpretation strategy for counterfactuals again as evaluating the consequent on the most similar models of the antecedent. But now it is not similarity between possible worlds or interpretation functions that counts, but similarity between the causal models that encode the relevant causal laws. Indeed, Pearl is able to establish an interesting result about the relation between his theory and particularly Lewis' (1973) logic of counterfactuals. Pearl gives an axiomatization of his theory that is sound and complete for causal models. This axiomatization consists of three axioms: composition, ef-

---

<sup>22</sup>There is an additional complication involved in Fine's Nixon example, that we have not discussed in section 5.3.2: the non-monotonicity of the involved law. A correct presentation of the law underlying the Nixon example should be along the following lines:  $button \wedge \neg abnormal \rightarrow holocaust$ , where *button* stands for the sentence that Nixon presses the button, *holocaust* for the sentence that a nuclear holocaust takes place, and *abnormal* represents a conjunction of abnormality conditions that may stop the law from working. One of these abnormality condition, then, would be that the circuit connecting the button with the nuclear weapons is not working. Reasoning with this law should involve negating all abnormality conditions as long as no information to the contrary is given. Neither the theory developed in Veltman (2005) nor Pearl's (2005) approach, nor the one that will be introduced later in this chapter employs non-monotonic reasoning with laws. We have to assume that  $\neg abnormal$  (and, hence, that the wire of the circuit is not cut) is given as a fact about the evaluation world. Then all three theories make correct predictions for the example. In the future these approaches should be extended with non-monotonic reasoning mechanisms to overcome this stipulation.

fectiveness, and recursiveness. He shows then that there is a translation of Lewis' counterfactual into the Pearl counterfactual and vice versa and that with respect to these translations (i) the axioms of the sound and complete axiomatization Lewis (1973) provides for his logic of counterfactuals are all fulfilled in Pearl's system, and (ii) that composition and effectiveness can be derived from Lewis' axioms. "In sum, for recursive models, the causal model framework does not add any restrictions to counterfactual statements beyond those imposed by Lewis' framework; the very general concept of closest world is sufficient." (Pearl, 2000: 242). Pearl also makes clear that the assumption of recursiveness is necessary. But this axiom actually carries the essence of causality: it claims that causal dependencies are not circular.<sup>23</sup> This shows that it is exactly causal asymmetry that is added by Pearl's system to the very general framework of Lewis.

While – as we have seen – we can understand Pearl's system as an instantiation of the similarity approach, so far we do not know the exact nature of the similarity relation that would give us Pearl's interpretation of *would have* conditionals. Of course, it would be nice to have a reformulation of Pearl's theory in terms of similarity. But the only thing we can say so far is that it looks as if the relevant similarity relation differs clearly from what has been proposed in premise semantics. According to Veltman (1985), (2000) and Kratzer (1989) the relevant class of laws holding in the evaluation world has to be kept constant between worlds that are compared and differences in similarity concerns differences in the interpretation of certain singular facts in these worlds. Pearl, on the contrary, proposes that in those models for the antecedent in which the consequent is evaluated certain laws of the evaluation model of the conditional do not hold.<sup>24</sup> These are the laws connecting the antecedent to its causal history. They are replaced by laws stating the truth of the antecedent, independent of all other facts and laws.

## Problems

Regardless of the advantages of Pearl's theory, it also has some drawbacks. They will be the topic of the present section. The idea that evaluating a counterfactual involves "surgery" on mechanisms that govern reality plays a central role in Pearl's (2000) approach. These surgeries are active manipulations of the causal structures coded in a causal model. The truth of the antecedent is forced on these structures. This is done by giving up the causal laws determining the value of the antecedent and introducing new ones that claim the antecedent to be true *by law*. It is very questionable that causal laws of this form exist independently of

---

<sup>23</sup>For a precise definition of the properties composition, effectiveness and recursiveness see Pearl (2000).

<sup>24</sup>The reader may stumble at this point and wonder: is this really what Pearl proposes? Indeed, it is. The process of intervention throws away entire laws. It is obvious that this cannot be true in general, we will come back to this point in the problem section below.

this technical context, and, hence, whether there is any causal aspect of the world that is best described with such a constant function. But this aspect of Pearl's (2000) theory also produces very concrete problems. First, it leads to some rather technical difficulties. Because by definition  $F$  can only describe the value of endogenous variables, we cannot evaluate a counterfactual relative to a causal model where any of the variables in the antecedent are background variables of the causal model. This made the introduction of the unmotivated proposition letter  $U$  necessary in the discussion of the circuit example and the Kennedy example. An elegant way to solve this problem would be to manipulate the interpretation function  $I$  instead of the causal model  $M$ . This would also help to get rid of another reason why background variables cannot be manipulated in the theory of Pearl. If we would allow  $F$  to be extended to background variables in order to cope with antecedents that contain background variables, then the definition of intervention would allow us to associate these variables with the formula  $\top$  or  $\perp$ . However, if it comes to the evaluation of the consequent, this function would not be relevant, because by the definition of truth (definition 5.5.3) we have to check the interpretation function  $I$  for the value of background variables and not the causal model  $M$ . Thus, the proposed extension of  $F$  would not have any effect. The problem with the idea of manipulating the interpretation function  $I$  instead of the causal model is that so far  $I$  only describes the interpretation of the background variables, while to make the antecedent true we also have to change the interpretation of endogenous variables. In the next section we will see how this problem can be solved.

There are also more theoretical objections to Pearl's approach to intervention. As said above, this notion of intervention manipulates what counts as causal law. To make some antecedent true, some laws are generally given up, other laws newly introduced. From an intuitive perspective this sounds obviously wrong. Intervention cannot give up a law in general, because this would mean to lose every of its instantiations in reality. This problem is not visible on the level of abstraction chosen by Pearl (2000), because on this level laws are no general statements. They do not universally quantify over any variable. But as soon as you introduce, for instance, time into the framework, you want the causal laws to generalize about this parameter, i.e. to hold for all times. Then a law may have more than one instantiation in a world, and Pearl's theory starts to make some very strange predictions. Assume, for instance, that yesterday Peter pushed Mary. This caused Mary to drop the glass she was holding. In consequence the glass broke. Assume, furthermore, exactly the same has happened again today. Then Pearl's theory, extended in a straightforward manner to times, predicts that the following *would have* conditional is true.

- (75) If Mary hadn't dropped her glass today, then the glass she was holding yesterday wouldn't have been broken.

The problem is that to make the antecedent true Pearl would in general give

up the causal relation between pushing and dropping things and in its place introduce the causal law that Mary (or anybody) never drops glasses (or anything).<sup>25</sup> This is so, because his notion of intervention manipulates the variable  $F$  of a causal model, the function that encodes the causal laws (see definition 5.5.5). In consequence, it is predicted that Mary would also not have dropped the glass yesterday. This is another argument pushing the point that intervention should not take place at the level of general laws or causal models.

Another group of problems of Pearl's theory that we want to discuss here concerns the way Pearl models causality in his approach. Pearl does not provide a theory of causality, at least in the sense of an explanation for causation. He rather provides a possibility to formally describe causal dependencies.<sup>26</sup> But that means that the causal models we assumed for the examples are purely stipulative, based on our intuitive understanding about which variables are connected by causal dependencies and which are not. Pearl does not provide us with tools that allow us to establish causal relationships. This is a common problem for theories on counterfactual reasoning that make reference to some set of laws – for instance, also the approach of Veltman (2005) stipulates the relevant laws for a concrete example. But it is still a problem of Pearl's approach.

Even though Pearl does not provide a theory of causation, some decisions he makes with his representation of causal laws are theoretical decisions about the nature of causality. One of these decisions is the assumption of what he calls the *determinism* of causal laws. The way his causal model works it is always the case that if you know the causes of a certain effect and you know the truth values these causes have, then you can always determine the truth value of the effect. Pearl gives some explicit arguments for why he makes this assumption: (i) determinism is more general than stochastic methods, (ii) determinism is more in tune with human intuitions, and (iii) determinism is needed to define concepts involving counterfactual reasoning. But he does not seem to be aware that he uses *determinism* in a very specific sense that differs from the general notion of determinism underlying the arguments (i) to (iii). There is some discussion in the literature on the point whether Pearl's determinism of causal laws is a reasonable assumption (see, for instance, Korb et al. 2005). In fact, there are also some examples for *would have* conditionals that appear to be problematic for Pearl's assumption of determinism. Consider again the Tichy example from section 5.4.1.

*Consider a man - call him Jones - who is possessed of the following disposition as regards wearing his hat. If the man on the news predicts bad weather, Mr Jones invariably wears his hat the next day. A*

---

<sup>25</sup>The exact formulation depends on how general the original law is formulated: whether it quantifies over all agents and all things that can be dropped etc.

<sup>26</sup>In the end, his primary goal is to answer the question how a robot should process causal information and not what causality is. He still is a student of Computer Science.

*weather forecast in favor of fine weather, on the other hand, affects him neither way: in this case he puts his hat on or leaves it on the peg, completely at random. Suppose, moreover, that yesterday bad weather was prognosed, so Jones is wearing his hat. ...*

The question is whether in this context you accept the conditional (76).

- (76) If the weather forecast had been in favor of fine weather, Jones would have been wearing his hat.

Intuitively, the answer is no. But does the approach of Pearl predicts these intuitions? The problem is to find a suitable causal model for the example. Presumably, there is a causal relation between the weather forecast and whether Mr. Jones is wearing his hat or not. But as far as the context tells us, this relation is not deterministic: while from the prediction of bad weather it follows that Mr. Jones wears his hat, nothing can be told about the location of the hat in case the weather is predicted to be fine. One may argue – and this is probably what Pearl would do – that there is a hidden, unknown causal factor  $X$  that, together with the weather forecast, determines whether Jones carries his hat or not. The relevant causal model could then be described as follows:  $M = \langle B, E, F \rangle$  with  $B = \{bad, X\}$ ,  $E = \{hat\}$ , and  $F(E) = bad \vee X$ , where *bad* is true if the weather is bad,  $X$  is the hidden cause, and *hat* is true if Mr. Jones carries his hat. Let us check what this model predicts for the conditional (76), formalized as  $\neg bad \succ hat$ . We first need again a dummy variable to be able to manipulate the weather conditions. Hence, the model becomes  $M' = \langle B', E', F' \rangle$  with  $B' = \{U, X\}$ ,  $E' = \{bad, hat\}$ ,  $F'(hat) = bad \vee X$ , and  $F'(bad) = U$ . Then we need an interpretation function for the background variables that fits the context described above. It is clear that  $U$  should be interpreted as true. But what about  $X$ ? The context tells us nothing about the hidden variable, let alone about its value. The best thing we can do is to assume incompetence of the interpreter about the value of  $X$  and distinguish two interpretation functions  $I_1$  and  $I_2$ : both map  $U$  on 1, but  $I_1$  maps  $X$  on 0, while  $I_2$  maps  $X$  on 1. We then calculate the truth value of  $\neg bad \succ hat$  with respect to  $M', I_1$  as well as  $M', I_2$ . The reader can check that the conditional comes out as false with respect to the first tuple, but as true with respect to the second. We can take  $I_1$  and  $I_2$  to characterize the belief state of some interpreter who does not know the truth value of the hidden cause. As our results show, such an interpreter would believe that the conditional (76) is false. That means that Pearl can account for the intuition that (76) is false as the result of incompetence about the value of some hidden cause. Whether this explanation is satisfying depends on how comfortable we are with this hidden variable. There seem to be no good reasons why we should load causality with the burden of the stipulation of hidden causes. If we can do without them, we should do so. But then we need to formulate Pearl with a non-deterministic conception

of causality.<sup>27</sup>

The most obvious shortcoming of Pearl's proposal is that it essentially reduces *would have* conditionals to those where a causal connection exists between antecedent and consequent. The theory proposed above predicts that if the consequent is not causally dependent on the antecedent (that means that there is no sequence of arrows leading from antecedent to consequent in the graph representing the relevant causal model), then a *would have* conditional is true if and only if its consequent is true.<sup>28</sup> Pearl does not think that this is a shortcoming of his approach, because he is convinced that there are only causal *would have* conditionals. Also Lewis (1979) claims that this is the result we want (for the normal resolution of similarity).<sup>29</sup> In the remainder of this section we will discuss various examples that question this prediction. There are true *would have* conditionals that are neither based on causal dependencies, nor reducible to the truth of the consequent. In the next section we will see how we can deal with these cases.

Let us start with a quite famous example from Kratzer (Kratzer 1989: 640) that Pearl's (2000) approach cannot handle.

*King Ludwig of Bavaria likes to spend his weekends at Leoni Castle. Whenever the Royal bavarian flag is up and the lights are on, the King is in the Castle. At the moment the lights are on, the flag is down, and the King is away. Suppose now counterfactually that the flag had been up.*

(77) If the flag had been up, then the King would have been in the castle.

This *would have* conditional is intuitively true. However, the approach defined above evaluates it as false because the only causal interpretation that seems reasonable is that King Ludwig's being in the castle causes the lights to be lit and the flag to be flown. Let *flag* be the proposition that the flag is up, *light* be the proposition that the light is on, and *king* the proposition that the King is in the castle. Then we are talking about causal laws  $F(\text{light}) = \text{king}$  and  $F(\text{flag}) = \text{king}$ . With respect to such a causal model flying a flag, hence, manipulating *flag*, cannot cause the King to come to the castle. Hence, (77) is predicted to be false. Intuitively, the problem seems to be that we work with

---

<sup>27</sup>Another respond to the example could be that there is no causality involved in the example. But a causal interpretation of Tichy's example is possible and Pearl should be able to account for it.

<sup>28</sup>The intervention with the antecedent has no effect on the truth-conditions of the consequent. After intervention the consequent has the same value as before.

<sup>29</sup>Lewis did not make the claim for causal dependencies but for temporal order: if the consequent of a *would have* conditional lies temporally before the antecedent, then the conditional is true if and only if the consequent is true (see Lewis 1979: 458).



the wrong laws: instead of  $F(\text{light}) = \text{king}$  and  $F(\text{flag}) = \text{king}$  this example is based on the law:  $F(\text{king}) = \text{light} \wedge \text{flag}$ . However, this is certainly not a causal law and, thus, not in the scope of Pearl's (2000) approach.<sup>30</sup>

The following conditionals make a similar point. They all represent plausible *would have* conditionals that are based on non-causal laws.

- (78) a. If the barometer had been low, then (probably) there would have been a storm.
- b. If there had been a storm, the barometer would have been low.
- c. A dice is thrown. Six comes up. If one had come up, then six would have been on the lower side.

The examples (78a) and (78b) are based on a law that describes a correlation, not a causal relationship. The acceptability of these *would have* conditional shows that we can have such *would have* conditionals. Example (78c) illustrates that analytical laws based on conventions can also be used for counterfactual reasoning. Even though the acceptability of these examples is arguably not as straightforward as for causal *would have* conditionals, they do occur. A theory of the semantics of *would have* conditionals cannot simply deny their existence.

A first idea on how to account for these examples may be to extend what counts as a law in Pearl's theory to other laws than causal ones. However, this will not do. Pearl's theory assumes that in certain circumstances causal laws can be broken. If we treat other laws on a par with causal laws, then we predict that in similar circumstances they can be broken as well. This is not true for analytical laws.

*If you are born in a year with an air pollution higher than X and are now older than 60, you run a high risk of getting a certain sort of pneumonia. Max was born in 1946, so he is 60 now. He checks and finds out that in 1946 the air pollution rate was lower than X, while between 1946 and 1953 the rate was always higher than X. He says:*

- (79) a. ??If I had been born some years later, I would run a high risk of getting pneumonia now.
- b. If I had been born some years later I wouldn't be 60 now.

As the low acceptability of (79a) shows<sup>31</sup>, we are not prepared to give up the analytical relation between the year of birth and the age of a person when

---

<sup>30</sup>A first idea one might have for how to account for this example is to use the notion of d-separation. There are, however, empirical problems with such an approach. For reasons of space we cannot discuss the details of such a proposal within the thesis.

<sup>31</sup>The conditional is acceptable without problems if you ignore the age-condition of the law. But this would mean to interpret the sentence based on a different law. We are not interested here in this reading.

thinking about what would have been the case, if Max were born in a different year. In such a situation the age changes as well, see (79b).

Fortunately, the class of laws that never can be broken appears to be separable from those that can. The first group consists of analytical logical laws, the second of causal laws and correlations.<sup>32</sup> One might, therefore, propose that in order to deal with the counterexamples of Pearl's theory we have to adapt the original proposal by treating different laws differently. Some can be broken, some cannot. But still we would be in trouble. Even for those laws that can be broken it seems always to be possible to come up with some examples where Pearl would predict them to be broken, while there exists a reading of the *would have* conditionals that behaves as if the law is valid. More precisely, we observe in these cases an ambiguity. Besides the reading predicted by Pearl (2000), there is also a reading available that is obeying the laws. Below, first two examples are given that illustrate this point for correlations. Afterwards, we will show that this even holds for *would have* conditionals based on causal relations. In all examples discussed below the a-sentence follows the predictions made by Pearl (hence, in the interpretation some laws are broken), while the b-conditional negates the a-sentence by taking exactly the same laws to be valid.

### Examples for correlations

*The state of the barometer correlates with the weather conditions. The barometer is low if and only if there is a storm.*

- (80) a. If the barometer hadn't been low, then we would have taken the boat and may all have drowned by now.  
 b. No. If the barometer hadn't been low, there wouldn't have been a storm.

*Every time when baby Simon is very thirsty he is sick the next day.*

- (81) a. If Simon hadn't been so thirsty yesterday, I wouldn't have noticed the symptoms so early.  
 b. No. If Simon hadn't been so thirsty yesterday, he wouldn't have been sick today.

### Examples for causal laws

*There is a court, an officer, two riflemen and a prisoner. If the court orders the execution, then the officer will give a signal to the riflemen. If the officer gives the signal to the riflemen, then the riflemen will*

---

<sup>32</sup>The two groups might not completely be characterized.

*shoot. If a rifleman shoots, then the prisoner will die. The court orders the execution. The officer gives the signal, the riflemen both shoot, and the prisoner dies.*

- (82) a. (Even) if rifleman A hadn't shot, the prisoner would have died.  
 b. No. If rifleman A hadn't shot then the court wouldn't have ordered the execution, the officer wouldn't have given the signal and the prisoner would still be alive.

*Ann sometimes goes to parties. Bob likes Ann very much but is not into the party scene. Hence, save for rare circumstances, Bob is at the party if and only if Ann is there. Carl tries to avoid contact with Ann since they broke up last month, but he really likes parties. Thus, save for rare occasions, Carl is at the party if and only if Ann is not at the party. Bob and Carl truly hate each other and almost always scuffle when they meet. Now consider the following discussion between two friends who did not go to the party but were called by Bob from his home. They observe that Ann must not be at the party, or Bob would be there instead of at home. But that must mean that Carl is at the party!*

- (83) a. If Bob were at the party, then Bob and Carl would surely scuffle.  
 b. No. If Bob was there, then Carl would not be there, because Ann would have been at the party.

This last example stems from Balke and Pearl (1994). The authors actually claim that the speaker of (83b) does not employ counterfactual reasoning but indicative reasoning and that this is reflected in the choice of '*was*' in place of '*were*'. They appear to argue that (83b) belongs syntactically to a different class of conditionals than (83a), but this is a difficult position to defend. It would mean that (83b) never can be understood in the same way as (83a) and the so called 'indicative' reading is not available for (83a). I think that both claims are empirically wrong.

The last two examples also illustrate another point. They both involve causal backtracking. All *would have* conditionals that employ backward causal reasoning are problematic for the approach of Pearl. In section 5.3.1, however, we have argued that such conditionals do exist. Hence, we have found another problem for Pearl's (2000) approach.

Let us conclude this discussion of the problems of Pearl's (2000) approach. There are two central findings of this section. First, even though Pearl's theory makes

very promising predictions for causal *would have* conditionals we have seen that it cannot deal with *would have* conditionals in general. Some restrictions are due to technical problems with the way the approach is set up. They may be solved without changing the basic ideas of Pearl (2000). Other restrictions are consequences of these basic ideas itself – as Pearl’s claim that only causal *would have* conditionals exist. To overcome them, more substantial changes of the approach are necessary. Second, the last examples have shown that *would have* conditionals are either ambiguous or context dependent. Otherwise it cannot be explained why given the same antecedent we can come to conclusions that contradict each other.

## 5.6 Two readings for conditionals

### 5.6.1 Motivation

In the following a new approach to the meaning of *would have* conditionals will be developed. It will combine Pearl’s causal theory for the meaning of *would have* conditionals with premise semantics. Central to the proposal stands the claim that two different readings for *would have* conditionals have to be distinguished: an *epistemic* and an *ontic* reading. The description that will be provided for these two readings will pick up on a fundamental distinction made in the philosophical/logical literature on the meaning of conditionals and belief revision. This is the distinction between *local* and *global* revision. Before we come to a detailed description of the two readings we will introduce the distinction between local and global revision and shortly review the related discussion.

To distinguish between two different paradigms for evaluating conditionals, and – what turns out to be closely related – two different strategies of belief revision has a long history in the literature on the similarity approach. Such a distinction is already inherent in Stalnaker’s (1968) famous paper *A theory on conditionals*. Stalnaker suggests that before dealing with the question of *what are the truth conditions of conditional statements?* we should start asking *how does one evaluate conditional statements?*. He proposes that our evaluation strategy follows a recipe formulated by Ramsey (1950).<sup>33</sup>

The Ramsey test condition (Stalnaker’s version)

“First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequence is then true.” (Stalnaker, 1980: 45)

---

<sup>33</sup>Stalnaker (1968) extends this recipe to cover also counterfactual conditionals.

This receipt suggest that the evaluation of conditional statements is based on belief revision. Stalnaker continues that we have to make a transition from this evaluation strategy to truth conditions of conditionals that explains why we use this method of evaluation. He proposes that for the truth conditions we can in principle use the same condition, but we have to replace ‘*stock of beliefs*’ by ‘*possible world*’.

Stalnaker’s truth conditions or conditionals (informally)

“Consider a possible world in which  $A$  is true, and which otherwise differs minimally from the actual world. “If  $A$ , then  $B$ ” is true (false) just in case  $B$  is true (false) in that possible world.” (Stalnaker, 1981a: 45)

According to this condition, truth of conditional statements can be described as revision of worlds instead of belief revision. Stalnaker (1981a) implicitly assumes that both are intrinsically connected: selecting for every world consistent with the beliefs the most similar antecedent world gives the same result as selecting the most similar belief state. However, it turns out that this is not true. Selecting minimally different belief states and selecting minimally different worlds describe different revision operations. Let us be more precise on this point. The standard description of the global revision process underlying revision as described by the Ramsey test condition can be defined as follows.<sup>34</sup>

#### Global Revision

Let  $K$  be a formula representing the information encoded in a belief state. Let  $\leq$  be a function that maps a formula  $K$  on a total pre-order  $\leq_K$  over models  $w$  and fulfills the following three conditions: (i) if  $w, w' \models K$ , then  $w \not\prec_K w'$ , (ii) if  $w \models K$  and  $w' \not\models K$ , then  $w <_K w'$ , and (iii) if  $K \leftrightarrow K'$ , then  $\leq_K = \leq_{K'}$ .

$$GlobalRev(K, \psi) = \{v \models \psi \mid \neg \exists u : u <_K v\}.$$

The global notion of revision selects those models of  $\psi$  that are closest to the knowledge base described by  $K$ . The belief conditions of conditionals could then be described as follows.

$$K \models A \succ B \text{ iff } GlobalRev(K, A) \models B$$

The local type of revision involved in Stalnaker’s truth conditions for conditionals can be characterized as follows.<sup>35</sup>

<sup>34</sup>This particular formulation is taken from Katsumono & Mendelzon (1991).

<sup>35</sup>The formulation follows Katsumono & Mendelzon (1991).

Local Revision for worlds

Let  $\leq$  be a function that maps a model  $w$  to a partial pre-order  $\leq_w$  over models  $w$  and fulfills the following condition: if  $w' \neq w$ , then  $w <_w w'$ .

$$LocRev(w, \psi) = \{v \models \psi \mid \neg \exists u : u <_w v\}.$$

Local revision can be extended to a description of belief conditions for conditionals by distribution over all worlds consistent with the beliefs of the relevant belief state.

Local Revision for belief states

$$K \models A \succ B \text{ iff } \bigcup_{w \models K} LocRev(w, A) \models B.$$

One can show that the logical properties of the global revision of a belief state  $K$  differ from those of local revision. While the first kind of revision is characterized by the famous AGM postulates<sup>36</sup>, the axiomatization of the local variant<sup>37</sup> shows some clear deviations. Crucial is the following difference. In case of global revision, if sentence  $\psi$  is consistent with the belief state, then revision gives you as a new belief state those models of the beliefs that satisfy  $\psi$ . In case of local revision, however, you can end up with worlds that are not consistent with the beliefs. This is to be expected: if one world consistent with the beliefs is inconsistent with  $\psi$ , then local revision cannot see whether there are other worlds in the belief state that  $\psi$  is consistent with. Kazumono & Mendelzon (1991) illustrate the difference between the two kinds of revision with the following example. Assume that we have a room with two objects in it, a book and a magazine. Suppose  $b$  means the book is on the floor, and  $m$  means the magazine is on the floor. Suppose, furthermore, that we are in a belief state where we believe that either the book is on the floor or the magazine is on the floor:  $(b \wedge \neg m) \vee (\neg b \wedge m)$ . Now we want to revise the belief state with the sentence  $b$  that the book is on the floor. Because this claim is consistent with our beliefs, global belief revision would predict that the result is the set of models of our belief state where  $b$  is true. But that means that from our new belief state we could conclude  $\neg m$ . Local revision does not warrant such a result. In this case we look for closest worlds for each model of our belief state separately. The possibility  $b \wedge \neg m$  is mapped to itself, because it already makes  $\neg m$  true. But why should  $\neg b \wedge m$  be mapped to  $b \wedge \neg m$  as well? Shouldn't similarity rather force us to keep  $\neg m$  constant in the process of revision? The only thing we can conclude from the local revision function is that if the belief state is consistent with the sentence it is to be revised with, then the result of revision will be consistent with the old belief state. Kazumono &

<sup>36</sup>See Alchourrón et al. (1985), for a proof of the claim see Katsumono & Mendelzon (1991).

<sup>37</sup>See Katsumoto & Mendelzon (1991).

Mendelzon (1992) suggest – following Keller & Winslett Wilkins (1985) – that this difference between global and local revision can be best understood as one between modifying a knowledge base when new information about a static world is obtained (in case of global revision) and bringing the knowledge base up to date when the world is changed (in case of local revision). This is interesting for us, because this description of local revision comes very close to Pearl’s characterization of the way we interpret *would have* conditionals. As explained in section 5.5, Pearl claims that the models relevant for testing the truth of the consequent are obtained by actively changing the model to make the antecedent true.

There exists a famous paradox concerning the global revision function. Global revision is intended to describe the conditions under which we believe a conditional *if A then B*. One might, thus, expect that the conditional is part of a belief state if and only if globally revising the belief state with *A* would lead to a belief state that contains the belief that *B*. This was actually Stalnaker’s conjecture for the probabilistic variant of global revision: conditionalization.

Stalnaker’s conjecture (Stalnaker, 1981a)

$$P(A > C) = P(C|A)$$

One can show, however, that this equivalence can only hold for trivial belief states.<sup>38</sup> Interestingly, the conjecture does hold for local belief revision.<sup>39</sup> The triviality result itself is not that relevant for our considerations. But very interesting is Stalnaker’s first reaction to it in Stalnaker (1981b). Here, he claims that indeed sometimes conditionalization should not be used to characterize the beliefs of an agent in a conditional. In particular, he observes that conditionalization leads to wrong predictions if applied to decision making. Assume that you want to calculate the expected utility of an action *A*. You distinguish a number of different outcomes  $B_1, \dots, B_n$  for this action to which you attach different utilities  $U(B_i)$ . You might then propose to calculate the expected utility of action *A* by taking the average of the utilities of the outcomes weighted by the conditional probability of the outcomes on performing action *A*.

$$EU(A) = \sum_{1 \leq i \leq n} P(B_i|A)U(B_i)$$

Stalnaker observes that sometimes this calculation does not provide a rational measure for which action should be chosen. This is in particular the case if *A* is evidentially relevant to the truth of  $B_i$ , but doing *A* has no causal influence on

---

<sup>38</sup>The original proof for the probabilistic statement is due to Lewis (1981), a proof for global belief revision as defined here can be found in Gärdenfors (1988).

<sup>39</sup>Lewis proves this claim for his probabilistic variant of local revision, *imaging*, in Lewis (1981). Grahne (1991) shows the same for local revision as defined here.

the outcome  $B_i$ . “Then  $P(B_i|A) > P(B_i)$ , but only an ostrich would count this as any sort of reason inclining one to bring it about that  $A$ . To do so would be to act as to change the evidence, knowing full well that one is in no way changing the facts for which the evidence is evidence.” (Stalnaker, 1981: 151). Stalnaker suggests that in this case one should use the probability of the conditional to calculate the expectations, which we know, by the triviality result, to diverge from the conditional probability  $P(B_i|A)$  sometimes.

$$EU(A) = \sum_{1 \leq i \leq n} P(A > B_i)U(B_i)$$

This would mean that – at least in the described contexts – the probability of the conditional depends as much on causal dependencies as on stochastic dependencies. However, it is not clear why this should be true for the Stalnaker conditional. Maybe he adopts here the position of Lewis (1973), who claims that causal dependence is to be explained as counterfactual dependence. Gibbart & Harper (1981) further develop the ideas Stalnaker (1981b) sketches. They simply assume an essentially causal meaning for conditionals.

The Gibbart & Harper Causal Paradigm for the meaning of conditionals  
 A counterfactual  $A \succ B_i$  is true either if (i)  $A$  brings about  $B_i$  or (ii)  $B_i$  would hold regardless of  $A$ .

This is by no means a fleshed-out theory for the meaning of *would have* conditionals, as the authors themselves observe. However, Gibbart & Harper (1981) sketch a particular similarity relation for local revision that they suppose to provide this meaning.

The Gibbart & Harper similarity approach to the meaning of conditionals  
 A counterfactual  $A \succ B_i$  is true in world  $w$  at time  $t$  if  $B_i$  holds in all worlds  $w'$  that fulfill the following conditions.

- $w'$  is like  $w$  before  $t$ ,
- the agent decides in  $w'$  to do  $A$  at  $t$ ,
- $w'$  obeys the physical laws from time  $t$  on, and
- $w'$  is maximally similar to  $w$  at  $t$  and the differences in the initial conditions at  $t$  should be entirely within the agents decision-making apparatus.

Gibbart & Harper (1981) argue that based on an interpretation of conditionals as given by the causal paradigm correct predictions are made for the expected utilities in the examples given by Stalnaker and that in all other cases the probability of this causal conditional equals the conditional probability.



Let us summarize the findings of this excursion into the philosophical/logical literature on the similarity approach. As we have seen, a distinction can be made between two different types of revising a belief state with new information: global revision and local revision. Both types of revision may be relevant for different applications. The question that concerns us here is in how far this distinction is also relevant for the interpretation of English *would have* conditionals. A number of authors have suggested that the difference between local and global revision explains the different meanings of indicative and subjunctive conditionals. Since Adam's (1970) Kennedy example it is commonly accepted that there is a semantic difference between the two types of conditionals.

- (84) a. If Oswald didn't shoot Kennedy, then somebody else did.  
       b. If Oswald hadn't shot Kennedy, then somebody else would have.

Intuitively, the sentences (84a) and (84b) seem to have different truth conditions. One can very well agree with the first while denying the second.<sup>40</sup> While Stalnaker wants to use local revision for all types of conditionals, Lewis thinks that this method is adequate for subjunctive conditionals like (84b) but not for indicative conditionals like (84a). Katsumono & Mendelzon (1992) suggest that local revision is more proper for describing the meaning of subjunctive conditionals (84b) than is global belief revision. Harper proposes that while global revision correctly captures the acceptability of indicative conditionals, for subjunctive conditionals one should rather adopt his causal paradigm for the evaluation of conditionals, which he and Gibbart (1981) suggest to be produced by a special instantiation of local revision. We quote here Harper (1981:19).

“The Ramsey test seems to accord quite well with the way we evaluate the acceptability of the indicative conditional (84a). For most of us, the claim that Kennedy was shot is a salient piece of what we take to be our accepted body of knowledge. When each of us hypothetically revises his body of knowledge to assume the antecedent that Oswald didn't shoot Kennedy, he retains this salient claim that Kennedy was shot. This, in turn forces high credence for the consequence that someone shot Kennedy. This Ramsey test reasoning seems to be the right account of the high credence most of us place in the indicative conditional (84a).

When we turn to the subjunctive conditional (84b) the causal paradigm is much more appropriate than Ramsey's test. Presumably, we would not accept (84b) unless we belief something like the following story:

---

<sup>40</sup>Adams (1970) and Skyrms (1976) both argue at length for modelling the meaning (or assertability) of indicative conditionals with subjective conditional probability. They propose that for subjunctive conditionals other (past) probability distributions be used, but this proposal is not relevant for our discussion here.

Other marksmen were in the position to shoot Kennedy if Oswald missed or failed to fire, or were there to shoot Kennedy regardless of Oswald.

This is exactly the kind of story that renders (84b) acceptable on the causally sensitive paradigm.”

We agree with Harper and many others in that global belief revision correctly models the interpretation of (84a). We also agree with Harper in that the second conditional (84b) is evaluated using local revision and that causality plays an important role for this interpretation. However, we will not follow Harper (1981) and others in proposing that this difference in interpretation is expressed by the use of different moods in both conditionals: the indicative mood in the first, and the subjunctive mood in the second. Instead, we will propose that for all conditionals, independent of mood, two readings have to be distinguished. First, there is what we call the epistemic reading of a conditional. The epistemic reading is based on belief revision. It is used for conditionals that make statements about what one would conclude upon learning that the antecedent is true. It reasons about what you would believe, if you learned – hypothetically – that the antecedent is true. From the epistemic reading we will distinguish an *ontic reading* of conditionals. This reading is applied if the conditional is interpreted as describing the consequences for the course of history it would have, if the antecedent were true. The ontic reading follows the observations made by Stalnaker (1981b) and Gibbart & Harper (1981) on the relevant interpretation of conditionals in the context of rational choice theory, particularly, with respect to the importance of causal dependencies for the evaluation of conditionals. We will, however, not assume, following Lewis, that the close relation between causal dependencies and the meaning of *would have* conditionals is a consequence of the fact that causal dependence is to be explained as counterfactual dependence. Instead, we will propose that causal dependencies go as input into the interpretation of conditionals. The formalization of the ontic reading will be highly inspired by the work of Pearl (2000). According to the ontic reading the antecedent is made true by intervention into the causal history of the evaluation world. But we will reformulate Pearl’s notion of intervention in terms of a local revision function.

Can we give linguistic evidence for the proposed ambiguity – besides the indirect evidence provided later that it can explain the data better than the other approaches discussed so far? One of the central conclusions drawn from the discussion of Pearl’s (2000) approach was that, indeed, conditionals allow for different readings. Somehow, any theory for the meaning of *would have* conditionals has to account for this observation. However, the existence of different readings still leaves different types of explanation available. Instead of proposing an ambiguity, one could have chosen for an underspecification approach. But such an

approach is not particularly fit to describe the existence of exactly two readings – which is what we seem to observe (see section 5.5.4). An interesting empirical argument for the distinction of exactly two different readings of conditionals comes from an observation we have made in section 5.3.1 where we discussed backtracking counterfactuals. There, we observed that the acceptability of the insertion of an additional modal *have to*<sup>41</sup> in the consequent of a *would have* conditional changes dependent on the reading applied to the conditional. The relevant difference seems to be correctly captured by the distinction between the epistemic reading and the ontic reading that we propose here. If this is correct, then *have to* insertion can be used as a test to distinguish between the two readings.<sup>42</sup>

The idea of distinguishing different readings for conditionals, in particular to distinguish between an ontic and an epistemic reading, is not new. Kaufmann (2005) argues for the same point with respect to indicative conditionals. Another reference is Kratzer (1981). Still, this claim is under heavy debate in the literature. Some philosophers, such as Rott (1999) and Veltman, argue vehemently that the epistemic reading does not exist. Others, for instance Morreau (1992), want to explain everything in terms of an epistemic reading. The debate between these two positions focusses on examples like the duchess example from Veltman (2005) (see section 5.4.3) or the similar Hamburger example from Hansson (1989).

*Suppose that one Sunday night you approach a small town of which you know that it has exactly two snackbars. Just before entering the town you meet a man eating a hamburger. You have good reason to accept the following indicative conditional:*

(85) *If snackbar A is closed, then snackbar B is open.*

*Suppose now that after entering the town, you see that A is in fact open. Would you now accept the following conditional?*

(86) *If snackbar A were closed, then snackbar B would be open.*

Opponents of the epistemic reading claim that the conditional (86) is simply unacceptable and the proposed epistemic reading, according to which the sentence comes out as true, does not exist. Notice, however, that Veltman accepts in Veltman (1985) a very similar example (see Veltman 1985: 217). Also Rott (1999) admits that there are subjunctive conditionals that obtain an epistemic reading, even though he holds that this is not possible for the context under discussion. Other philosophers have no doubt about the acceptability of such a

---

<sup>41</sup>Rott (1999) makes the same observation for *must* insertion.

<sup>42</sup>In the course of this thesis we will not make any proposal for the interpretation of the extra modal *have to* in the consequent of conditionals or provide an explanation for why it improves on the acceptability of epistemic conditionals, while leading to unacceptable ontic conditionals. This is left to future work.

conditional in the described situation, such as, for instance, Morreau (1992). In sum, in light of the data it seems difficult to deny the existence of an epistemic reading of *would have* conditionals. On the other hand, philosophers like Morreau (1992) that argue for an epistemic-only approach to the meaning of *would have* conditionals have no easy position either. Not only is it difficult to explain examples like Lifschitz' circuit example based on epistemic reasoning, any unique-meaning approach to *would have* conditionals has problems to account for the debate around the intuitions concerning examples like (86). We will propose that (86) is false according to its dominant ontic reading, but true according to its deficient epistemic reading.

### 5.6.2 The epistemic reading

We will start with developing a formal description of the epistemic reading of *would have* conditionals. As explained above, the idea is that the epistemic reading is about what an interpreter infers upon (hypothetically) learning that the antecedent is true. Hence, it is based on belief revision. To get an idea of how to describe the relevant notion of belief revision let us turn to a standard example of the epistemic reading just mentioned: the duchess example of Veltman (2005). We want the epistemic reading to describe the interpretation of (88) according to which the conditional is true in the given context (Veltman 2005: 174).

*'The duchess has been murdered, and you are supposed to find the murderer. At some point only the butler and the gardener are left as suspects. At this point you believe*

*(87) If the butler did not kill her, the gardener did.*

*Still, somewhat later – after you found out convincing evidence showing that the butler did it, and that the gardener had nothing to do with it – you get in a state, in which you will reject the sentence*

*(88) If the butler had not killed her, the gardener would have.'*

As the quote shows, Veltman (2005) claims that in this context (88) is false. Many people disagree with Veltman on this point and claim that there is a reading of (88) in the provided context according to which the conditional comes out as true. This alternative reading we will describe as the epistemic reading of (88). To be precise, also Veltman sees the possibility of an epistemic reading, but he thinks that this interpretation can hardly ever be communicated. According to him (2005) the epistemic reading is calculated as follows: "In the epistemic case implicit reference is made to some previous epistemic state, in this example the state you were in when only two suspects were left. Thinking back one can say that if it had not been the butler, it would have been the gardener." (Veltman,

2005: 174). This description of the epistemic interpretation strategy cannot be correct. This can easily be seen with the following example where the order in which the crucial information is learned by the inspector is reversed, but still intuitively the conditional (88) can be interpreted as true.<sup>43</sup>

### **The duchess example – a variation**

*Last night the duchess was murdered in her sleep. You are supposed to find the murderer. Soon after the investigations start the lab calls. They have found fingerprints of the butler all over the crime scene. You interrogate the butler and he confesses. At this point you believe that the butler did it, and that the gardener had nothing to do with it. Somewhat later the lab calls again. They have checked all the locks of the house. None is broken. There are only two persons besides the duchess that have keys for the house: the butler and the gardener. Now, you believe:*

(88) *If the butler had not killed her, the gardener would have.*

In this variation of the duchess example there is no past belief state at which the inspector believed the indicative conditional *If the butler did not kill her, the gardener did*. Thus, it cannot be reference to past belief states that is responsible for the reading of (88) in the two contexts according to which the sentence is true. We have to find a different explanation. What we need in order to account for the example is that when giving up the belief that the butler killed the duchess, the interpreter still has to hold on to the belief that either the butler or the gardener did it. Then, assuming that the butler did not kill her will lead to a belief state where it is true that the gardener did it. This intuitive account for the duchess example can easily be formalized using premise semantics for belief states. We assume that every belief state  $K$  is characterized by (i) a set of facts true in it – the set selected by the premise function, or the *basis* of the belief state using Veltman's (2005) words – and (ii) the general laws assumed to hold in this state. The revision of the belief state  $K$  with a sentence  $\psi$  is then, roughly, the set of worlds that model  $\psi$  and maximal subsets of the premises consistent with the general laws and  $\psi$ . So far, the approach is standard premise semantics. But somehow, we have to explain why for the given example the premises contain the fact that either the butler or the gardener killed the duchess. To account for this, we propose that belief revision is sensitive to a distinction between facts we learn from observation, hence, facts we have some sort of independent external

---

<sup>43</sup>Readers that have difficulties to get the intended reading should try the following variant with *have to* in the consequent.

(89) If the butler had not killed her, the gardener would have to have done it.

evidence for, and facts we derive from this input by general laws. The premises of a belief state are proposed to be given by the first set of facts: facts we have external evidence for. When revising our beliefs, we try to keep, in addition to all laws, as much as we can of these facts. The next section will make this idea formally precise.

### 5.6.2.1 Formalization

Let us start by laying down the basics of the framework in which we will formalize the epistemic reading. The formal language we will use is a propositional language. It is closed under the operators  $\neg$  and  $\wedge$ . We add a second binary operator  $>$ . Sentences  $\psi > \phi$  are to represent epistemic *would have* conditionals. We do not allow for iterated uses of the operator  $>$ .<sup>44</sup>

#### 5.6.1. DEFINITION. (Language)

Let  $\mathcal{P}$  be a set of proposition letters. The language  $\mathcal{L}^0$  is the closure of  $\mathcal{P}$  under negation and conjunction. The language  $\mathcal{L}^>$  is the union of  $\mathcal{L}^0$  with the set of expressions  $\psi > \phi$  for  $\psi, \phi \in \mathcal{L}^0$ .

Now, we have to define the model with respect to which we interpret expressions of  $\mathcal{L}^>$ . We assume a possible worlds approach to truth. The truth conditions of sentences in  $\mathcal{L}^0$  are defined following standard lines. As proposed in the introduction, the truth of sentences  $\psi > \phi$  is based on belief revision with the antecedent  $\psi$ . That means that for their truth conditions we access belief states. Leaving the definition of a belief state open for the moment, we can describe the model for the language  $\mathcal{L}^>$  we need as follows.

#### 5.6.2. DEFINITION. (Worlds and models)

A *possible world* for  $\mathcal{L}^>$  is an interpretation function  $w : \mathcal{P} \longrightarrow \{0, 1\}$ . A *model*  $M$  for  $\mathcal{L}^>$  is a tuple  $\langle W, K \rangle$ , where  $W$  is a set of possible worlds and  $K$  is a function mapping worlds to belief states. For  $\psi \in \mathcal{L}^>$  we write  $M, w \models \psi$  in case  $\psi$  is true with respect to  $M$  and  $w$ .  $\llbracket \psi \rrbracket^M$  is the set of possible worlds  $w \in W$  such that  $M, w \models \psi$ .

It is a common practice to model belief states in possible world approaches by accessibility relations between possible worlds. This will not do for our purposes. We need more information from the representation of a belief state than just the set of possible worlds consistent with the beliefs. More particularly, we need to know which general laws are taken to hold with respect to a belief state and what is the set of facts of the belief state for which the agent has independent external evidence. This leads to the following definition. We use a standard

---

<sup>44</sup>To be precise, the definition of the language  $\mathcal{L}$  allows for no embedded occurrences of the conditional connector  $>$ , but this restriction could be easily lifted without problems.

notion of satisfaction, according to which a set of sentences  $A$  is satisfiable in a set of possible worlds  $W' \subseteq W$  of a model  $M = \langle W, K \rangle$ , if there is some element of  $W'$  that (together with  $M$ ) makes all elements of  $A$  true.

### 5.6.3. DEFINITION. (Belief state)

Let  $M = \langle W, K \rangle$  be a model for the language  $\mathcal{L}^>$ . A *belief state* is a tuple  $\langle \mathcal{B}, U \rangle$ , where  $\mathcal{B} \subseteq \mathcal{L}^0$  is a finite set of sentences and  $U \subseteq W$  a set of possible worlds such that  $\mathcal{B}$  is satisfiable in  $U$ .  $\mathcal{B}$  is called the *basis* of the belief state,  $U$  its *universe*.  $\llbracket \langle \mathcal{B}, U \rangle \rrbracket^M$  is the set  $\{w \in W \mid w \in U \text{ \& } M, w \models B\}$ .

Following the line of thought developed above, the basis  $\mathcal{B}$  of a belief state is the set of sentences for which the agent of the belief state has independent external evidence. Intuitively, a possible world belongs to the universe  $U$  of a belief state  $\langle \mathcal{B}, U \rangle$ , if it makes all general laws true the agent of the belief state takes to be valid.<sup>45</sup>

One may wonder whether this tuple  $\langle \mathcal{B}, U \rangle$  also gives a conceptually complete characterization of a belief state. Is it not possible that the agent of the belief state engages other beliefs besides those derivable from  $\mathcal{B}$  and  $U$ ? We will assume that this is not the case.

#### Assumption

There are no beliefs that the agent of a belief state entertains that are not derivable from the set of facts the agent takes to be given by external evidence and the general laws he considers to be valid.

Based on the given characterization of belief states we can now formulate truth conditions for epistemic conditionals. In this definition we make reference to a global revision function *Learn* that still has to be defined. Notice, that according to this approach the truth of a conditional  $\psi > \phi$  does not depend at all on the non-modal facts of the evaluation world, but only on the belief state this world is associated with.

### 5.6.4. DEFINITION. (The epistemic reading of *would have* conditionals)

Let  $M = \langle W, K \rangle$  be a model for  $\mathcal{L}^>$ ,  $w$  an element of  $W$ , and  $\psi, \phi \in \mathcal{L}^0$ . The conditional  $\psi > \phi$  is true with respect to  $M$  and  $w$  if  $\psi$  is true on  $\text{Learn}_M(K(w), \psi)$ :

$$M, w \models \psi > \phi \text{ iff } \text{Learn}_M(K(w), \psi) \models \phi.$$

---

<sup>45</sup>One might wonder whether the complexity of this notion of belief state is really necessary. Why not follow Veltman (2005) and define belief states as tuples  $\langle F, U \rangle$  where  $U$  corresponds to our  $U$  and  $F$  is the set of worlds consistent with the beliefs of the agent of the belief state? However, we cannot derive  $B$  from this belief state – for instance, as the minimal set of sentences that describes  $F$  given  $U$ . The reason is that the agent of the belief state may have independent external evidence for sentences that (relative to  $U$ ) logically depend on each other. This is illustrated by the duchess example (88). At the point where the speaker utters (88) he has independent external evidence for the sentence *Either the butler or the gardener killed the duchess* and the sentence *The butler killed the duchess*. On learning that the butler did not do it, he is prepared to give up the second sentence. But, as (88) shows, he still sticks to the first.

That was easy. But now we have to say how the function *Learn* is defined. If *Learn* is a function that describes belief revision, then it should be a function from a belief state  $\langle \mathcal{B}, U \rangle$  and a sentence  $\psi$  to a new belief state  $\langle \mathcal{B}', U' \rangle$  – the belief state you obtain when revising you old beliefs with the information  $\psi$ . How to model belief revision? First, we assume that revision cannot change what counts as the laws. If *Learn* is applied to a sentence  $\psi$  that is inconsistent with the laws, then revision breaks down. This is an assumption often entertained, but also one that does not seem to be empirically correct. You can learn information that stands in conflict with some law you believe to be valid. Intuitively in such a situation the critical law is given up. We simplify matters here, because we do not want to get involved with this additional possibility of belief change.

Also for the revision of the basis of a belief state we follow standard lines. We try to keep as many of the elements of  $\mathcal{B}$  as possible – that means as satisfiable together with  $\psi$  in  $U$ . The well-known problem with this standard approach to belief revision is that in general there exist many such maximal subsets of  $\mathcal{B}$  satisfiable with  $\psi$  in  $U$ . Which of them is describing the basis of the new belief state after revision? The classical answer to this question is to take as  $\mathcal{B}'$   $\psi$  together with the intersection of all these candidates for the revised belief state.<sup>46</sup> But this approach to belief revision has some rather unwanted consequences. Taking the intersection of the maximal subsets of  $\mathcal{B}$  satisfiable together with  $\psi$  intuitively makes you lose too many of your old beliefs.<sup>47</sup> For illustration, consider the set  $\mathcal{B} = \{\neg A, \neg B\}$ . We want to revise a belief state with this basis with the sentence  $A \vee B$ . There are two maximal subsets of  $\mathcal{B}$  that are satisfiable together with  $\psi$ <sup>48</sup>:  $\{\neg A\}$  and  $\{\neg B\}$ . Defining the revised basis as the intersection of  $\{\neg A, A \vee B\}$  and  $\{\neg B, A \vee B\}$  would give use  $\mathcal{B}' = \{A \vee B\}$ . Thus, not only do we give up both beliefs  $\neg A$  and  $\neg B$ , we are even prepared to consider it possible that both beliefs were false at the same time. It seems more convincing to model belief revision in a way that one sticks at least to the belief  $\neg A \vee \neg B$ . But how to predict this within a framework where belief states are modeled using set of sentences?

Hansson (1989) proposed a very interesting solution for this problem that still allows us to define *Learn* as function from belief states to belief states. But because we are mainly interested in the meaning of *would have* conditionals and not belief revision, we will take another way out that is more adequate for our application. This is a solution inherent in the premise semantics approach to the meaning of conditionals developed in Veltman (1976). We assume that all maximal subsets of  $\mathcal{B}$  satisfiable, together with  $\psi$ , in  $U$  are, when extended with

---

<sup>46</sup>This is known as belief revision based on full meet contraction, see Alchourrón & Makinson (1981).

<sup>47</sup>This problem cannot be solved to full satisfaction using the concept of partial meet contraction, discussed in Alchourrón et al. (1985).

<sup>48</sup>We consider  $U$  in this case to be unrestricted: the set of all interpretation functions for  $A$  and  $B$ .



$\psi$ , candidates for the revised belief state. But as far the meaning of an epistemic conditional is concerned, we do not have to know which of the candidates is actually chosen. We propose that an epistemic conditional  $A > C$  is true with respect to the belief state  $\langle \mathcal{B}, U \rangle$ , if  $C$  holds in all the potential results of revising  $\langle \mathcal{B}, U \rangle$  with  $A$ .

Let us compare this approach with the one we outlined first: we take as the set of worlds on which the consequent is checked the worlds  $\llbracket \langle \mathcal{B}', U \rangle \rrbracket^M$ , where  $\mathcal{B}'$  is the intersection of all maximal satisfiable subsets plus the antecedent. We want to calculate the truth conditions of a conditional  $(A \vee B) > C$  with respect to a belief state  $\langle \mathcal{B}, U \rangle$  with  $\mathcal{B} = \{\neg A, \neg B\}$  and  $U$  the set of all interpretation functions for  $\mathcal{P} = \{A, B, C\}$ . We already know how the revised belief state  $\langle \mathcal{B}', U' \rangle$  looks in case we model belief revision by intersecting all maximal satisfiable subsets of  $\mathcal{B}$ :  $U' = U$  and  $\mathcal{B}' = \{A \vee B\}$ . The relevant worlds are marked by the dashed area in the left picture of figure 5.5 below. If we take instead the interpretation strategy of premise semantics and check the consequent in all belief states you get by defining the second approach to truth conditions of epistemic conditionals just as discussed as the union of  $\{A \vee B\}$  with the maximal subsets of  $B$  satisfiable together with  $A \vee B$  in  $U$ , the  $C$  has to be true in the worlds marked in the right picture of figure 5.5. As we can see, the second description of the truth conditions of *would have* conditionals is truly weaker. The consequent has to be true on a smaller set of worlds.

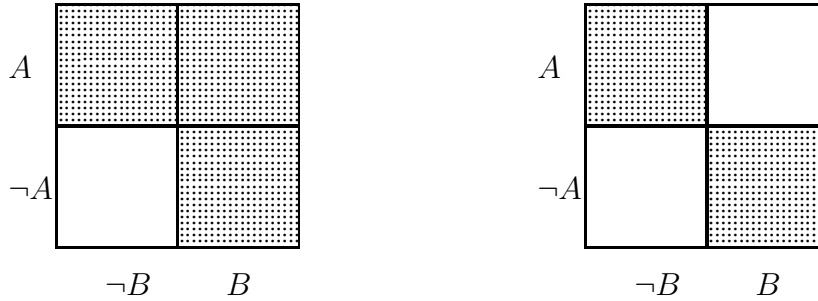


Figure 5.5: Two strategies of belief revision

Below we give a definition of the function *Learn* that implements this second approach to the truth conditions of epistemic conditionals just discussed. First we introduce an order between possible worlds  $w$  that compares how many elements of the basis of the relevant belief state are true in  $w$ . Then, the function *Learn*, applied to a belief state  $\langle \mathcal{B}, U \rangle$  and a sentence  $\psi$  is defined as selecting minimal elements with respect to this order.

##### 5.6.5. DEFINITION. (Order induced by a belief state)

Let  $M = \langle W, K \rangle$  be a model for  $\mathcal{L}^>$  and  $\langle \mathcal{B}, U \rangle$  a belief state for  $M$ . We define a partial order on the elements of  $U$  as follows.

$$\forall u_1, u_2 \in U : u_1 \leq^{\langle \mathcal{B}, U \rangle} u_2 \text{ iff } \{\varphi \in \mathcal{B} \mid M, u_1 \models \varphi\} \supseteq \{\varphi \in \mathcal{B} \mid M, u_2 \models \varphi\}$$

**5.6.6. DEFINITION.** (The minimality operator)

Let  $D$  be any domain of objects and  $\leq$  an order on  $D$ . The minimality operator  $Min$  is defined as follows:

$$Min(\leq, D) = \{d \in D \mid \neg \exists d' \in D : d' < d\}$$

**5.6.7. DEFINITION.** (Belief revision)

Let  $M = \langle W, K \rangle$  be a model for  $\mathcal{L}^>$ ,  $w$  an element of  $W$  with  $K(w) = \langle \mathcal{B}, U \rangle$ , and  $\psi$  an element of  $\mathcal{L}^0$ .  $Learn_M(\langle \mathcal{B}, U \rangle, \psi)$ , is defined as follows:

$$Learn_M(\langle \mathcal{B}, U \rangle, \psi) = Min(\leq^{\langle \mathcal{B}, U \rangle}, \llbracket \psi \rrbracket^M \cap U).^{49}$$

**5.6.2.2 Discussion of the epistemic reading**

The approach proposed above can account for the reading of the duchess example of Veltman (2005) according to which the conditional (88) is true, the similar reading of the Hamburger example of Hansson (1989), as well as the King Ludwig example of Kratzer (1989). Let us illustrate this in some more detail. We start with the duchess example repeated above as example (88), (see page 129). Let *butl* be the proposition that the butler killed the duchess, *gard* the proposition that the gardener killed her, and *dead* the proposition that the duchess is dead. In the context where (88) is situated there is independent external evidence for *butl*, for *dead* and for *butl*  $\vee$  *gard*. Hence,  $\mathcal{B} = \{\textit{butl}, \textit{dead}, \textit{butl} \vee \textit{gard}\}$ . The function  $Learn_M(\langle \mathcal{B}, U \rangle, \neg \textit{butl})$  returns the set of worlds in  $U$  where  $\neg \textit{butl}$  is true and a maximal subset of  $\mathcal{B}$  that together with  $\neg \textit{butl}$  are satisfiable in  $U$ . There is only one maximal subset of  $\mathcal{B}$  with this property:  $\{\textit{dead}, \textit{butl} \vee \textit{gard}\}$ . Hence,  $Learn_M(\langle \mathcal{B}, U \rangle, \neg \textit{butl}) = \{w \in U \mid M, w \models \{\textit{butl} \vee \textit{gard}, \textit{dead}, \neg \textit{butl}\}\}$ . On this set it is true that the gardener is the murderer. Thus, (88) is predicted to be true according to its epistemic reading.

The proposed approach can also account for Kratzer's (1989) King Ludwig example (77) repeated here as (90).

---

<sup>49</sup>Because  $\mathcal{B}$  is a finite set of sentences the maxima will always exist. Lewis (1973) introduces an elegant way to deal with the possibility that maxima (or minima) do not exist for similarity approaches to the meaning of conditionals (i.e. approaches that select models for the antecedent on which the consequent is checked by looking for minima/maxima with respect to certain orders). The evaluation conditions are reformulated as follows: for every model of the antecedent there exists some smaller/larger model for antecedent and consequent. We could implement this solution in the present framework. We refrain from doing so because it would unnecessarily complicate the approach.

*King Ludwig of Bavaria likes to spend his weekends at Leoni Castle. Whenever the Royal bavarian flag is up and the lights are on, the King is in the Castle. At the moment the lights are on, the flag is down, and the King is away. Suppose now counterfactually that the flag had been up.*

(90) *If the flag had been up, then the King would have been in the castle.*

Let *flag* be the proposition that the flag is up, *light* the proposition that the light is on and *king* the proposition that King Ludwig is in the castle. We propose that the normal interpretation of the context is such that the interpreter assumes the speaker of the conditional to be looking at the castle from a distance. Hence, he has external evidence for  $\neg flag$  and *light*, but not for  $\neg king$ . The latter is derived from the other two by a general law of the form  $(flag \wedge light) \leftrightarrow king$ . Hence,  $\mathcal{B} = \{\neg flag, light\}$  and  $U$  is the set of possible worlds where the law  $(flag \wedge light) \leftrightarrow king$  holds. The function  $Learn_M(\langle \mathcal{B}, U \rangle, flag)$  will keep the basis-fact *light* true when giving up  $\neg flag$  of  $\mathcal{B}$ . From the resulting set of possible worlds we can then derive that the King is in the castle. Thus, (90) is predicted to be true.

In the same way one can also account for another example in Veltman (2005). The author observes that in the context given below (91) is not true (Veltman 2005: 178).

*“Consider the case of three sisters who own just one bed, large enough for two of them but too small for all three. Every night at least one of them has to sleep on the floor. Whenever Ann sleeps in the bed and Billie sleeps in the bed, Carol sleeps on the floor. At the moment Billie is sleeping in bed, Ann is sleeping on the floor, and Carol is sleeping in bed. Suppose now counterfactually that Ann had been in bed ...*

(91) *Well, in that case Carol would be sleeping on the floor.”*

The present approach predicts the conditional (91) to be false, because in the most straightforward interpretation there is external evidence for the location of each of the three sisters. In consequence, among the closest worlds where Ann is in the bed are worlds where Carol is on the floor, and there are also worlds where Billy is on the floor. A referee of Veltmann (2005) suggests a variation of the context where (91) is intuitively true (Veltman 2005: 178).

*“Suppose Carol is invisible. Suppose further that you are a proud parent of Ann, Billie and Carol, and before you go to bed you go and check on the kids. As described in the original version, Ann is on the floor, Billie is in bed and Carol (obviously) is also in bed. Now you turn to your spouse and comment:*

(92) If Ann had been in bed, Carol would have been on the floor.”

Our proposal predicts (92) to be true in its epistemic reading. The reason is that this time there is no external evidence for the location of Carol. Hence, the fact that Carol is in the bed is not part of the basis  $\mathcal{B}$  of the relevant belief state. It is easily given up to maintain the order-relevant basis-fact that Billy is in the bed.

According to the epistemic reading of *would have* conditionals described here causal backtracking is possible. As the system is set up, belief revision will always bring us to worlds where all laws still hold. In particular, all causal laws have to hold. That means that if the antecedent describes the effect of some causal law and the law says that this effect can only hold when a certain cause occurred, then *Learn* will select worlds where indeed the cause did occur. So, backward reasoning using causal laws is possible.

We have already said at various places above that the formal description of the epistemic reading given here is an application of premise semantics for conditionals (see Veltman 1976 and Kratzer 1979, 1981a). The well-known problem of premise semantics is to answer the question *what are the premises?*. The main contribution of the present work is the answer provided for this question: the premises are the basis facts of the relevant belief state; the facts the agent of the belief state has independent external evidence for. But also this idea is not entirely new. For instance, it has been suggested in the literature on belief revision that the facts to which the revision operation applies should only be a subset of all the facts believed by some agent: “The intuitive processes [of contraction and revision, the author] themselves, contrary to casual impression, are never really applied to theories as a whole, but rather to more or less clearly identified bases of them.” (Alchourrón & Makinson 1982: 21). Even the particular interpretation of this basis we have chosen here has been proposed earlier. For instance, Hansson (1989) suggested in reaction to the Hamburger example (cited in section 5.6.1 above) something similar: “In general, a case can be made for representing beliefs by sets that contain only the primary beliefs that have independent grounds. We (hopefully) believe the logical consequences of our primary beliefs, but these logical consequences should be subject only to exactly those changes (revisions or contractions) that follow from the changes of the primary beliefs.” (Hansson, 1989: 118). A similar suggestion can be found in Veltman (2005) in reaction to Kratzer’s (1989) King Ludwig example (see (90), page 136): “Clearly there is an important difference between, on the one hand, the king’s presence and, on the other hand, the light being on and the flag being up; the latter serve as external signs for the otherwise invisible occurrence of the former.” (Veltman 2005: 178). Veltman suggests that this may be of relevance for the interpretation of the example. The present approach makes these intuitive ideas of Hansson and Veltman precise by defining the basis of a belief state, to which the operation

of revision applies, as the set of facts for which the agent of the belief state has independent external evidence.

An important issue within the literature on belief revision is whether a given description of belief revision fulfills certain postulates assumed to be minimal requirements for rational belief change. As we defined the function *Learn* it is, in a strict sense, not a function of belief revision, because it does not return an object of the type of a belief state. Nevertheless, it is possible to investigate, whether the output of *Learn* fulfills the postulates. The four statements listed below are the famous AGM postulates applied to the present framework (Alchourrón et al., 1985).

The AGM-postulates of belief revision

Let  $M = \langle W, K \rangle$  be a model for  $\mathcal{L}^>$ ,  $\langle \mathcal{B}, U \rangle$  a belief state of  $M$ , and  $\psi, \phi \in \mathcal{L}^0$ .

- (R\*1) For any sentence  $\psi \in \mathcal{L}^0$ ,  $\text{Learn}_M(\langle \mathcal{B}, U \rangle, \psi) \subseteq \llbracket \psi \rrbracket^M$ .
- (R\*2) If  $\llbracket \psi \rrbracket^M \neq \emptyset$ , then  $\text{Learn}_M(\langle \mathcal{B}, U \rangle, \psi) \neq \emptyset$ .
- (R\*3) If  $\llbracket \langle \mathcal{B}, U \rangle \rrbracket^M \cap \llbracket \psi \rrbracket \neq \emptyset$ , then  $\text{Learn}_M(\langle \mathcal{B}, U \rangle, \psi) = \llbracket \langle \mathcal{B}, U \rangle \rrbracket^M \cap \llbracket \psi \rrbracket^M$ .
- (R\*4) If  $\text{Learn}_M(\langle \mathcal{B}, U \rangle, \psi) \cap \llbracket \phi \rrbracket^M \neq \emptyset$ ,  
then  $\text{Learn}_M(\langle \mathcal{B}, U \rangle, \psi \wedge \phi) = \text{Learn}_M(\langle \mathcal{B}, U \rangle, \psi) \cap \llbracket \phi \rrbracket^M$ .

It is easy to see that (R\*1) holds for the belief revision function *Learn* defined here. (R\*2) holds as long as  $\psi$  is consistent with  $U$ , i.e. does not stand in conflict with general laws. Given that we do not want to deal with the possibility that  $\psi$  is contradicting laws, this limitation is nothing we have to worry about so far. (R\*3) is true for *Learn*. If  $\psi$  is true in some worlds in  $\llbracket \langle \mathcal{B}, U \rangle \rrbracket^M$ , then these worlds are selected by the belief revision with formula  $\psi$ . The left-to-right direction of (R\*4) is also valid for the function *Learn*.<sup>50</sup> The other direction, however, does not hold.<sup>51</sup> But actually it is very controversial whether a system of belief revision, or at least counterfactual reasoning, should have this property. There have been counterexamples brought forward against the validity of this principle. For instance, it has been claimed that the following reasoning based on the right-to-left direction of (R\*4) is not sound.

*Premise 1: If Verdi and Satie had been compatriots, Satie and Bizet might have been compatriots.*

*Premise 2: If Verdi and Satie had been compatriots, Bizet would have been French.*

---

<sup>50</sup>The proof is left to the reader.

<sup>51</sup>For a counterexample see the appendix.

*Conclusion: If both Verdi and Satie, and Satie and Bizet had been compatriots, Bizet would have been French.*

Let  $A$  be the sentence *Verdi and Satie are compatriots*,  $B$  the sentence *Satie and Bizet are compatriots*, and  $C$  the sentence *Bizet is French*. The example can then be reformulated as the following reasoning scheme.

*Premise 1: If had been A, might have been B.*

*Premise 2: If had been A, would have been C.*

*Conclusion: If  $A \wedge B$ , would have been C.*

If you assume that a *might* conditional is true if the sentence in scope of *might* in the consequent is satisfiable on the result of revision with the antecedent, then premise 1 instantiates the assumption of postulate (R\*4):  $Learn_M(\langle \mathcal{B}, U \rangle, A) \cap \llbracket B \rrbracket^M \neq \emptyset$ . According to our interpretation rule for epistemic *would have* conditionals (definition 5.6.4), the conditional in premise 2 is true if the following holds:  $Learn_M(\langle \mathcal{B}, U \rangle, A) \subseteq \llbracket C \rrbracket^M$ . For the conclusion of the reasoning scheme we obtain the truth conditions  $Learn_M(\langle \mathcal{B}, U \rangle, A \wedge B) \subseteq \llbracket C \rrbracket^M$ . We see that if the right-to-left direction of (R\*4) was valid, then the reasoning quoted should be sound. This is, however, not confirmed by intuitions. The following weaker and less controversial version of (R\*4) does hold for *Learn*.<sup>52</sup>

(R\*4w) If  $Learn_M(\langle \mathcal{B}, U \rangle, \psi) \subseteq \llbracket \phi \rrbracket^M$ ,  
then  $Learn_M(\langle \mathcal{B}, U \rangle, \psi \wedge \phi) = Learn_M(\langle \mathcal{B}, U \rangle, \psi) \cap \llbracket \phi \rrbracket^M$ .

### 5.6.3 The ontic reading

In this section we will describe the ontic reading of *would have* conditionals. This reading works by hypothetically changing the facts about the evaluation world of the conditional. The goal is to make the antecedent true. In the resulting worlds it is checked whether the consequent of the conditionals is true as well. (Compare this to the epistemic reading, where the beliefs of some agent about the evaluation world are changed.) The changes applied to the evaluation world are not at all subtle: the course of history is broken and the truth of the antecedent forced on reality. This process can lead to the violation of causal laws. The proposed interpretation strategy for ontic conditionals can be compared to the execution of hypothetical and idealized experiments: you implement certain starting conditions in a closed system and then investigate the consequences. An ideal experiment forces the starting conditions on the system without affecting any condition that is assumed to be causally independent of the starting conditions. This is exactly how we interpret ontic *would have* conditionals.

---

<sup>52</sup>The proof is left to the reader.

How is this idea formalized? Contrary to anything Lewis would have approved, the approach presented here uses as a starting point a representation of the general laws that are assumed to hold, in particular the causal laws. The interpretation itself is implemented along standard lines: a *would have* conditional is considered true in world  $w$ , if on the output of a local revision function applied to  $w$  and the antecedent of the conditional the consequent is true. The local revision function is formulated using premise semantics. The general structure of this revision function is similar to the global revision used for the epistemic reading. But this time we distinguish three sets of premises: besides the general laws and a set of basic facts, also the facts derivable from the basis and the laws will be relevant for the order. Furthermore, we will not demand that all laws have to be kept in the process of revision. Causal laws may be broken. Finally, we use a different set of basic facts. In case of the ontic reading the basis describes, so to say, the initial conditions of the evaluation world. For the definition causal dependencies will play a crucial role.

### 5.6.3.1 Formalization

As for the epistemic reading we define a formal language by adding to a standard propositional language an additional binary connective. We use a different connective this time:  $\gg$ . Sentences  $\psi \gg \phi$  are to express ontic readings of *would have* conditionals with antecedent  $\psi$  and consequent  $\phi$ .

### 5.6.8. DEFINITION. (Language)

Let  $\mathcal{P}$  be a finite set of proposition letters. The language  $\mathcal{L}^0$  is the closure of  $\mathcal{P}$  under conjunction and negation.  $\mathcal{L}^\gg$  is the union of  $\mathcal{L}^0$  with the set of expressions  $\psi \gg \phi$  where  $\psi, \phi \in \mathcal{L}^0$ .

The next thing we need is the model with respect to which the language will be interpreted. Meaning will again be defined with respect to a set of possible worlds, let's call it  $U$ , the *universe* of a model. But this time we will be a bit more concrete on the interpretation assigned to this set of possible worlds.  $U$  is understood as the set of worlds consistent with what is assumed to be the laws, more particularly, the logical and analytical laws. We explicitly do not demand the causal laws to restrict  $U$ . We need to keep track of these laws independently. One reason is that we need more information from causal laws than is represented by just letting them restrict the domain of possible worlds. We need to have access to their 'direction'. That means we have to be able to distinguish between cause and effect. Second, a crucial property of the ontic reading of *would have* conditionals is, according to the present proposal, that it can break causal laws. That means that for their evaluation we have to allow for worlds that violate causal laws. These considerations motivate the following definition of a model (the exact definition of a causal structure will be given afterwards).

**5.6.9. DEFINITION.** (Worlds and models)

A *possible world* for  $\mathcal{L}^\gg$  is an interpretation function  $w : \mathcal{P} \longrightarrow \{0, 1\}$ . A *model*  $M$  for the language  $\mathcal{L}^\gg$  is a tuple  $\langle C, U \rangle$ , where  $C = \langle B, E, F \rangle$  is a causal structure and  $U$ , the *universe*, is a set of worlds. For  $\psi \in \mathcal{L}^\gg$  we write  $M, w \models \psi$  in case  $\psi$  is true with respect to  $M$  and  $w$ .  $\llbracket \psi \rrbracket^M$  denotes the set of worlds  $w \in U$  such that  $M, w \models \psi$ .

A partial interpretation function  $i$  of  $\mathcal{P}$  *follows*  $U$  if  $\exists w \in U : i \subseteq w$ .  $I$  is the set of all partial interpretation functions of  $\mathcal{P}$  that follow  $U$ .

A *causal structure* will be defined closely related to Pearl's definition of a causal model. However, we will not use the exact definition given when the approach of Pearl was introduced. The function  $F$  will be defined in a different way, to get rid of Pearl's restriction to deterministic causal laws. In definition 5.5.1 of section 5.5.2  $F$  associated every endogenous variable  $Y$  with a formula  $\phi_Y$ . The truth value of  $Y$  was then defined as the truth value of this formula (see definition 5.5.3). But this way, as soon as the value of each proposition letter occurring in  $\phi_Y$  is defined, the value of  $Y$  is determined as well. To loosen this connection we will now associate an endogenous variable  $Y$  with (i) an n-tuple  $Z_Y$  of proposition letters – those proposition letters the value of  $Y$  directly depends on<sup>53</sup> – and, to describe the dependency, (ii) a *partial* truth function  $f_Y$  from the value of these letters to the value of  $Y$ . It is crucial here that  $f_Y$  may be partial. Hence, for some valuation of the elements of  $Z_Y$  the value of  $Y$  may not be defined. This accounts for the possibility of non-deterministic causal laws.

**5.6.10. DEFINITION.** (Causal structure)

Let  $\mathcal{P}$  be a finite set of proposition letters and  $\mathcal{L}^0$  the language you obtain when closing  $\mathcal{P}$  under negation and conjunction. A *causal structure* for  $\mathcal{L}^\gg$  is a triple  $C = \langle B, E, F \rangle$ , where

- i.  $B \subseteq \mathcal{P}$  are called *background* variables;
- ii.  $E = \mathcal{P} - B$  are called *endogenous* variables; and
- iii.  $F$  is a function mapping elements  $Y$  of  $E = \mathcal{P} - B$  to tuples  $\langle Z_Y, f_Y \rangle$ , where  $Z_Y$  is an n-tuple of elements of  $\mathcal{P}$  and  $f_Y$  a partial truth function  $f_Y : \{0, 1\}^n \longrightarrow \{0, 1\}$ .  $F$  is rooted in  $B$ .

The definition of the notion of rootedness remains in principle unchanged.

**5.6.11. DEFINITION.** (Rootedness)

Let  $B \subseteq \mathcal{P}$  be a set of proposition letters and  $F$  a function mapping elements  $Y$  of  $E = \mathcal{P} - B$  to tuples  $\langle Z_Y, f_Y \rangle$ , where  $Z_Y$  is an n-tuple of elements of  $\mathcal{P}$  and  $f_Y$  a partial truth function  $f_Y : \{0, 1\}^n \longrightarrow \{0, 1\}$ . We introduce the relation  $R_F$

---

<sup>53</sup>These are the proposition letters that in Pearl's approach occurred in the formula  $\phi_Y$ .



that holds between two proposition letters  $X, Y \in \mathcal{P}$  if  $X$  occurs in  $Z_Y$ . Let  $R_F^T$  be the transitive closure of  $R_F$ . The  $R_F$ -minima of a letter  $X \in \mathcal{P}$ ,  $Min_{R_F}(X)$ , are defined as follows:

$$MIN_{R_F}(X) = \{Y \in \mathcal{P} \mid R_F^T(Y, X) \ \& \ \neg \exists Z \in \mathcal{P} : R_F^T(Z, Y)\}.$$

We say that  $F$  is *rooted* in  $B$  if  $R_F^T$  is acyclic and  $\forall X \in \mathcal{P} - B : Min_{R_F}(X) \subseteq B$ .

The best way to represent the function  $F$  is by a set of truth tables that list in the top row the elements of  $Z_Y$  and  $Y$  and below the output of  $f_Y$  for every combination of values for  $Z_Y$ . If for some valuation the function is undefined, we will put a star in the respective cell for  $Y$ . To illustrate the working of the definition of a causal structure, let us provide a causal structure  $C = \langle B, E, F \rangle$  for the Tichy example that we used to criticize Pearl's assumption of causal determinism.

*Consider a man - call him Jones - who is possessed of the following disposition as regards wearing his hat. If the man on the news predicts bad weather, Mr Jones invariably wears his hat the next day. A weather forecast in favor of fine weather, on the other hand, affects him neither way: in this case he puts his hat on or leaves it on the peg, completely at random. Suppose, moreover, that yesterday bad weather was prognosed, so Jones is wearing his hat. ... .*

Let *bad* be the proposition letter expressing that the weather forecast is in favor of bad weather, and *hat* a proposition letter expressing that Mr. Jones is wearing his hat. The law that we want to capture with the causal structure  $C$  is that the state of the weather causally influences Mr. Jones conditions for wearing his hat in that if the weather is bad, he wears his hat. The most straightforward way to go is to take *bad* to be the background variable and *hat* to be an endogenous variable. Then, we chose  $Z_{hat} = \langle bad \rangle$  and for  $f_{hat}$  the partial function  $f_{hat} : \{0, 1\} \longrightarrow \{0, 1\}$  that maps 1 to 1 and is undefined for 0. The complete causal structure of the Tichy example plus graph is given in figure 5.6.

Now that we have defined the language as well as the model, we can provide truth conditions for sentences  $\psi$  of  $\mathcal{L}^{\gg}$ . Truth for sentences in  $\mathcal{L}^0$  is defined according to standard lines. The truth conditions we still have to provide are those of ontic conditionals  $\psi \gg \phi$ . The basic setup of this definition is not very surprising. We follow the local revision approach to the meaning of conditional sentences. A conditional with antecedent  $\psi$  and consequent  $\phi$  is said to be true with respect to a model  $M$  and a world  $w$ , if the consequent is true with respect to those  $w'$  you obtain by applying the local revision function *Intervene* to  $w$  and the antecedent  $\psi$ .

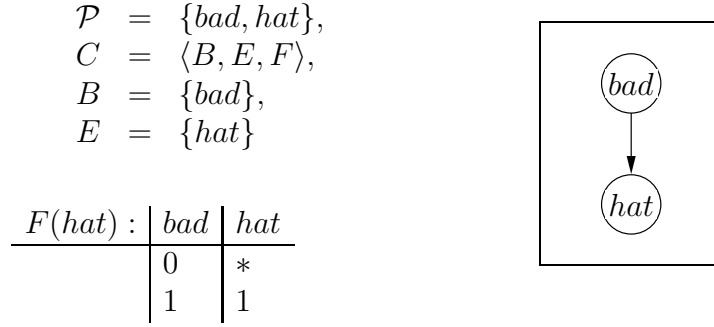


Figure 5.6: A causal structure for the Tichy Example

**5.6.12. DEFINITION.** (The ontic reading of *would have* conditionals)

Let  $M = \langle C, U \rangle$  be a model for  $\mathcal{L}^{\gg}$  and  $w \in U$  a possible world. For  $\psi, \phi \in \mathcal{L}^0$  we define that the sentence  $\psi \gg \phi$  is true with respect to  $M$  and  $w$  if  $Intervene_M(w, \psi)$  entails  $\phi$ :

$$M, w \models \psi \gg \phi \text{ iff } Intervene_M(w, \psi) \models \phi.$$

The revision function works – as is common for similarity approaches – by selecting worlds least different from the evaluation world  $w$  according to some order. The order is defined using premise semantics. As mentioned in the introduction to the ontic reading, we distinguish three different sets of premises, (i) the set of analytical/logical laws taken to hold in the evaluation world, (ii) the basis of a world, and (iii) the facts derivable from the laws and the basis. Crucial is how we define the basis. The basis will be described as the set of facts of a world from which, together with the laws, all other facts of  $w$  can be derived. So far this sounds as if we are using the same notion of a basis as does Veltman (2005), discussed in section 5.4.2. However, we will interpret what can be derived by laws differently. More particularly, we will demand that if a causal law is applied in the process of derivation, then the derivation has to follow the direction of the causal law, i.e. reason from cause to effect and not the other way around. Such a different notion of derivation leads, of course, to a different set of basis facts than used in Veltman (2005). In our case, the basis provides, roughly, the initial conditions of a world when everything started. On the first view, one might think that such a basis is simply given by the interpretation of the background variables of the relevant causal structure. But as said before, to model the ontic reading of conditionals we want to allow for worlds where causal laws can be violated. This means that not in all worlds is the interpretation of all endogenous variables correctly predicted by the evaluation of the background variables and the causal structure. To give a complete description of the facts of such a world the law-violating facts have to be part of the basis as well. This complicates the definition

of a basis. Below, we introduce first the *law closure* of a partial interpretation function  $i$ . This is the extension of  $i$  with the interpretation of proposition letters that can be derived by laws from  $i$ . Crucial here is that only derivations from causes to effects are allowed. This is realized very simply by prohibiting backward derivation starting from any endogenous variables interpreted by  $i$ .

**5.6.13. DEFINITION.** (Law closure)

Let  $M = \langle C, U \rangle$  be a model for  $\mathcal{L}^{\gg}$  and  $i \in I$  a partial interpretation of  $\mathcal{P}$ . The *law closure*  $\bar{i}$  of  $i$  is the minimal  $i'$  in  $I$  fulfilling the following conditions.<sup>54</sup>

- (i)  $i \subseteq i'$ ,
- (ii)  $i' = \bigcap \{w \in U \mid i' \subseteq w\}$ ,
- (iii) for all  $P \in E$  with  $Z_P = \langle P_1, \dots, P_n \rangle$  such that  $i(P)$  is undefined the following holds: if for all  $k \in \{1, \dots, n\}$ :  $i'(P_k)$  is defined and  $f_P(i'(P_1), \dots, i'(P_n))$  is defined, then  $i'(P)$  is defined and  $f_P(i'(P_1), \dots, i'(P_n)) = i'(P)$ ,

The following simple fact makes sure that this definition is well-formed.

**5.6.14. FACT.** Let  $M = \langle C, U \rangle$  be a model for  $\mathcal{L}^{\gg}$  and  $i \in I$  a partial interpretation of  $\mathcal{P}$ . The law closure  $\bar{i}$  of  $i$  is uniquely defined.<sup>55</sup>

The basis of a world  $w$  will be defined as the union of all smallest subsets of  $w$  (thus, partial interpretation functions) for which  $w$  is the law closure.

**5.6.15. DEFINITION.** (Basis)

Let  $M = \langle C, U \rangle$  be a model for  $\mathcal{L}^{\gg}$ . The *basis*  $b_w$  of a world  $w \in U$  is the union of all interpretation functions  $b \in I$  that fulfill the following two conditions: (i)  $b \subseteq w \subseteq \bar{b}$  and (ii)  $\neg \exists b' : b' \subseteq w \subseteq \bar{b'} \ \& \ b' \subset b$ .

Based on this notion of the basis of a world we can now define the result of the local revision function *Intervene* applied to a world  $w$  and a formula  $\psi$ . *Intervene* selects those worlds making  $\psi$  true that (i) have a least different basis from the evaluation world  $w$ , and (ii) have the greatest similarity with respect to derivable facts. We will, thus, introduce two orders on possible worlds, one comparing similarity with respect to bases and one comparing similarity with respect to derivable facts. The way the first order is defined differs to some extent from standard lines of premise semantics. We do not only demand that the overlap with the basis  $b_w$  of the evaluation world  $w$  is maximal, but also that the difference, calculated by set-subtraction, is minimal. This gives a more sensitive

---

<sup>54</sup>The relevant order with respect to which the minimum is calculated is set-inclusion between interpretation functions.

<sup>55</sup>For a proof see the appendix.

measure of similarity between bases than only selecting for maximal overlap. The same extension is not needed for the order of derivable facts. Here, we only look for maximal overlap. Measuring the differences does not make so much sense on this level – as least as long as we understand possible worlds as completely defined interpretation functions. This will change in the next chapter. Then, we will formulate the second order parallel to the first. But so far the following definitions of the orders are sufficient.

**5.6.16. DEFINITION.** (The orders)

Let  $M = \langle C, U \rangle$  be a model for  $\mathcal{L}^\gg$ ,  $w \in U$  a possible world. We define a function  $\leq_1$  that maps a world  $w$  on the following order: for  $w_1, w_2 \in U$  :  $w_1 \leq_1^w w_2$  iff (i)  $b_{w_1} \cap b_w \supseteq b_{w_2} \cap b_w$ , and (ii) if  $b_{w_1} \cap b_w = b_{w_2} \cap b_w$ , then  $b_{w_1} - b_w \subseteq b_{w_2} - b_w$ . Furthermore, we define a function  $\leq_2$  that maps a world  $w$  to the following order: for  $w_1, w_2 \in U$  :  $w_1 \leq_2^w w_2$  iff  $(w_1 - b_{w_1}) \cap (w - b_w) \supseteq (w_2 - b_{w_2}) \cap (w - b_w)$ .

The revision of world  $w$  with formula  $\psi$  is now determined as the set of minimal worlds with respect to these two orders. For the selection of the minima the order in which the orders are applied is important. We first select maximally similar bases and only in a second step pick out the worlds that show maximal similarity with respect to the derivable facts. This is an expression of the greater relevance of the basis for similarity than of facts that can be derived from the basis and general laws. But in contrast to Veltman (2005) we do not claim that this last set of facts is of no relevance at all.

**5.6.17. DEFINITION.** (Intervention)

Let  $M = \langle C, U \rangle$  be a model for  $\mathcal{L}^\gg$ ,  $w \in U$  a possible world. The *Intervene*-revision of  $w$  with a formula  $\psi \in \mathcal{L}^0$ ,  $Intervene_M(w, \psi)$ , is now defined as follows

$$Intervene_M(w, \psi) = Min(\leq_2^w, Min(\leq_1^w, \llbracket \psi \rrbracket^M)).$$

**5.6.18. FACT.** Let  $M = \langle C, U \rangle$  be a model for  $\mathcal{L}^\gg$ ,  $w \in U$  a possible world, and  $\psi$  an element of  $\mathcal{L}^0$ . For all  $w, w' \in U$  it holds that if  $w =_I^w w'$ , then  $w = w'$ .<sup>56</sup>

Assume that the antecedent is just some proposition letter  $P$ , part of the endogenous variables of the relevant causal structure, that is not an element of  $b$ , nor is its negation. We are interested in the outcome of  $Intervene_M(w, P)$  for some world  $w$ . If  $P$  is true in  $w$ , then  $Intervene_M(w, P) = \{w\}$ . There is only one basis for every world and this basis can have only one causal closure:  $w$  itself. Thus,  $w$  is the unique closest world to  $w$ . If  $P$  is false according to  $w$ , then the bases of the worlds selected by  $Intervene_M(w, P)$  contain  $P$  as additional basis fact (maybe some other changes will be necessary to comply with the analytical/logical laws of the relevant model). This normally means that the world generated by this basis violates causal laws in its interpretation of  $P$ .<sup>57</sup>

<sup>56</sup>For a proof see the appendix.

<sup>57</sup>This is not the case if the causal laws describing the interpretation of  $P$  is non-deterministic.

For illustration we will calculate whether the conditional (93) is true in the Tichy context on page 142.

- (93) If the weather forecast had been in favor of fine weather, Jones would have been wearing his hat.

We can define a model  $M = \langle C, U \rangle$  for the Tichy context by using as  $C$  the causal structure described in figure 5.6 and take  $U$  to be the set of all complete interpretation functions that can be defined based on  $\mathcal{P} = \{bad, hat\}$ . That means that we assume for the example no additional restrictions by analytical or logical laws. The ontic reading of the *would have* conditional (93) is formalized as the sentence  $\neg bad \gg hat$ . The world  $w$  with respect to which this sentence is going to be interpreted is  $\{bad, hat\}$ . The question we want to answer is whether  $M, w \models \neg bad \gg hat$ . This is true in case  $Intervene_M(w, \neg bad) \models hat$ . So, we have to calculate  $Intervene_M(w, \neg bad)$ . Figure 5.7 plots in the table on the left all worlds in  $U$  ( $w = w_4$ ). The basis of each of these worlds is marked by boxes around the elements of the interpretation function the basis is defined for. The picture to the right of the table shows how the worlds are related by the order  $\leq_1^w$ : an arrow points from world  $w_1$  to world  $w_2$  if and only if  $w_1 <_1^w w_2$ . As the figure illustrates,  $Min(\leq_1^w, \llbracket \neg bad \rrbracket^M) = \{w_1, w_2\}$ . For  $(w_1 - b_{w_1}) \cap (w - b_w)$  we calculate  $\emptyset$ . The same result we obtain for  $(w_2 - b_{w_2}) \cap (w - b_w)$ . Hence,  $w_1 =_2^w w_2$ . From this we can conclude  $Intervene_M(w, \neg bad) = \{w_1, w_2\}$ . On this set it does not hold that Jones wears his hat. Thus, the theory correctly predicts that the conditional (93) is not true.

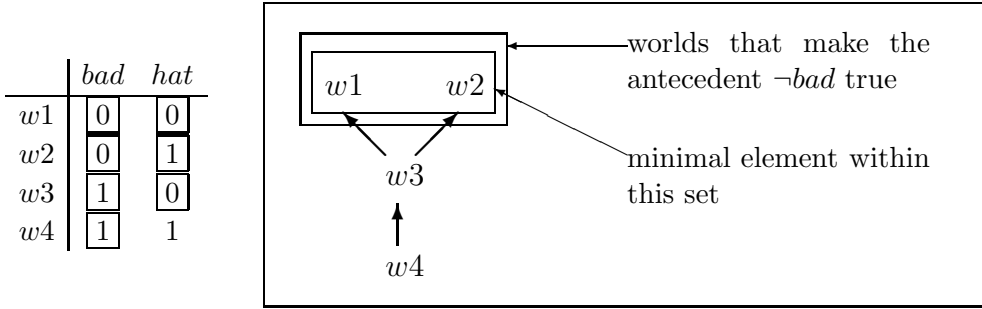


Figure 5.7: Minimal worlds for the Tichy example

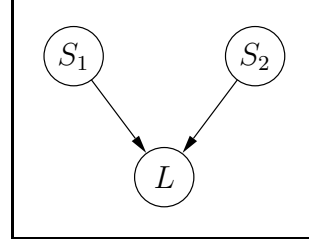
### 5.6.3.2 Discussion of the ontic reading

Let us illustrate how this approach to the ontic reading of *would have* conditionals accounts for some more examples. We start with Lifschitz' circuit example.

*Suppose there is a circuit such that the light is on exactly when both switches are in the same position (up or not up). At the moment switch one is down, switch two is up and the lamp is out. Now consider the following would have conditional:*

(94) If switch one had been up, the lamp would have been on.

$\mathcal{P} = \{S_1, S_2, L\},$   
 $M = \langle \langle B, E, F \rangle, U \rangle,$   
 $U = \text{all interpretation functions for } \mathcal{P},$   
 $B = \{S_1, S_2\},$   
 $E = \{L\}.$



$F(L) :$	$S_1$	$S_2$	$L$
	0	0	1
	0	1	0
	1	0	0
	1	1	1

Figure 5.8: A model for the Lifschitz Example

The model described by this context is given in figure 5.8.  $S_1$  stands for switch one being up,  $S_2$  for switch two being up, and  $L$  for the lamp being on. The causal structure is identical to the first causal model proposed for the Lifschitz example in section 5.5.2, figure 5.1. In contrast to the approach of Pearl we do not need turn the proposition letter  $S_1$  into a endogenous variable here. The present approach can handle antecedents that contain background variables. To see whether according to the ontic reading (94) is true, we have to calculate whether  $M, w \models S_1 \gg L$ , where  $w$  interprets  $S_1$  as false,  $S_2$  as true, and  $L$  as false. This is the case, if  $Intervene_M(w, S_1) \models L$ . Figure 5.9 lists all worlds of  $U$  together with their basis (marked by boxes around the relevant entries in the truth table), and on the right the way these worlds are related by the order  $<_1^w$ . As one can see,  $Min(\leq_1^w, \llbracket S_1 \rrbracket^M) = \{w_8\}$  and, hence,  $Intervene_M(w, S_1) = \{w_8\}$ . In  $w_8$  it is true that the lamp is on. Hence, the approach predicts correctly that on its ontic reading (94) is true.

Next we check the predictions made for the Kennedy example.

*Assume that there was a big conspiracy to kill Kennedy. They planned the assassination attempt of Oswald, but also a whole sequence of other attempts carried out by different people. Just by accident Oswald was the first one to succeed in killing Kennedy.*

(95) If Oswald hadn't killed Kennedy, someone else would have.

The model described by this context is given in figure 5.10.  $K_1$  represents that Oswald kills Kennedy,  $K_2$  that somebody else kills Kennedy, and  $D$  that

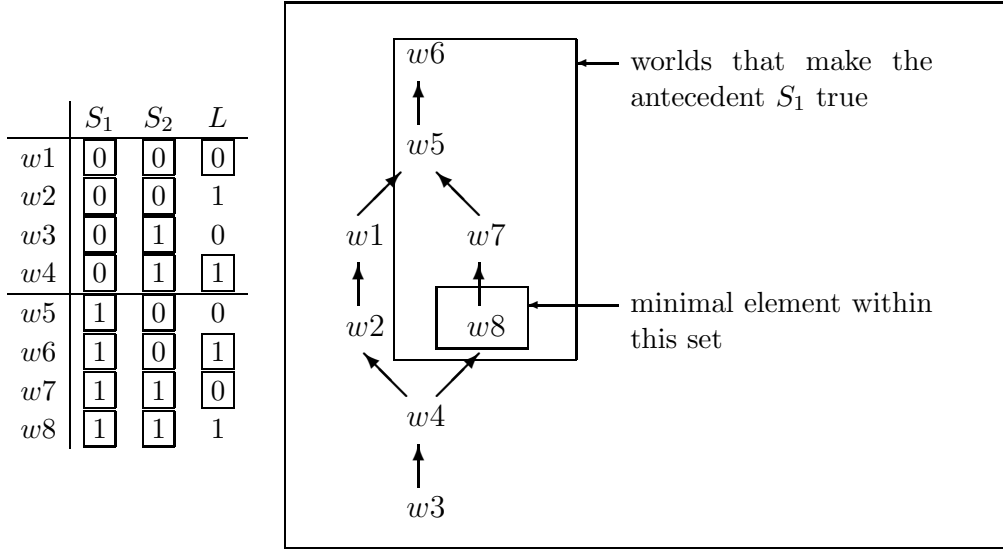
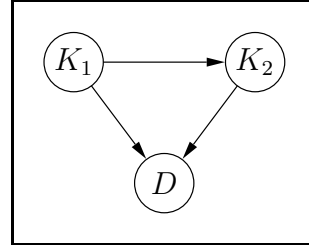


Figure 5.9: Minimal worlds for Lifschitz' circuit example

$$\begin{aligned}
\mathcal{P} &= \{K_1, K_2, D\}, \\
M &= \langle \langle B, E, F \rangle, U \rangle, \\
U &= \text{all interpretation functions for } \mathcal{P}, \\
B &= \{K_1\}, \\
E &= \{K_2, D\}, \\
F(K_2) &= \neg K_1, \\
F(D) &= K_1 \vee K_2
\end{aligned}$$



$F(K_2) :$	$K_1$	$K_2$
	0	1
	1	0

$F(L) :$	$K_1$	$K_2$	$D$
	0	0	0
	0	1	1
	1	0	1
	1	1	1

Figure 5.10: A model for the Kennedy Example

Kennedy is dead. The causal structure assumed here is the same as used in section 5.5.3, except that we do not need the extra variable  $U$  to turn  $K_1$  into an endogenous variable. To see whether according to the ontic reading (95) is true, we have to calculate whether  $\text{Intervene}_M(w, \neg K_1) \models K_2$ , where  $w$  maps  $K_1$  on 1,  $K_2$  on 0, and  $D$  on 1. Again, figure 5.11 lists the elements of  $U$  with their basis and the way they are related by the order  $<_1^w$ . As the reader can see,  $\text{Intervene}_M(w, \neg K_1) = \{w_4\}$ . On  $w_4$  Oswald did not shoot Kennedy, but somebody else did. Hence, this approach predicts that on its ontic reading (95)

is true.

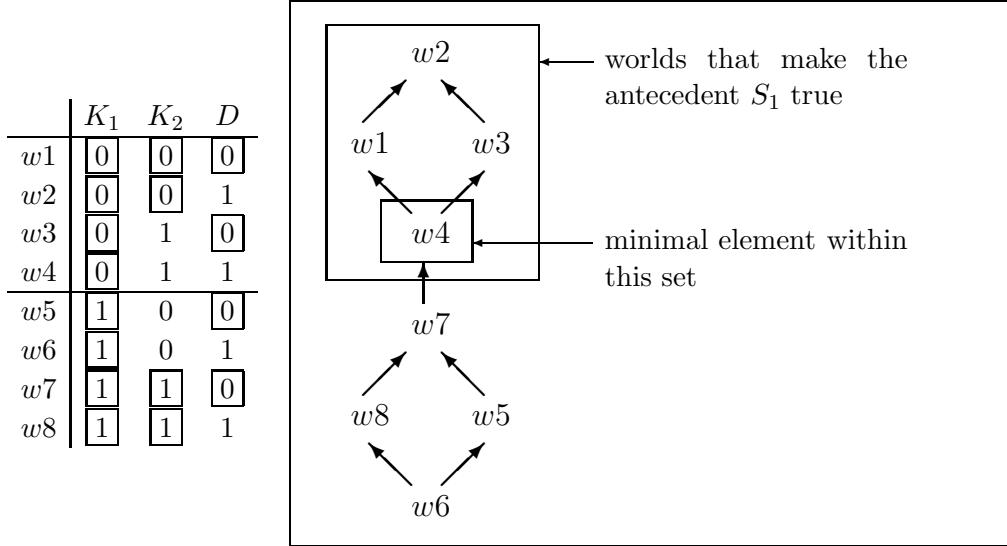


Figure 5.11: Minimal worlds for the Kennedy example

Finally, let us have a look again at the famous shooting squad example.

*There is a court, an officer, two riflemen and a prisoner. If the court orders execution then the officer will give a signal to the riflemen. If the officer gives the signal to the riflemen, then the riflemen will shoot. If a rifleman shoots, then the prisoner will die. The court orders the execution. the officer gives the signal, the riflemen both shoot, and the prisoner dies.*

(96) (Even) If rifleman A hadn't shot, the prisoner would have died.

The model described by this context is given in figure 5.4 in section 5.5.3. We will not repeat it here. Remember that  $C$  stands for the court orders the execution,  $O$  for the officer gives the signal,  $R_1$  for rifleman 1 shoots,  $R_2$  for rifleman 2 shoots, and  $P$  for the prisoner dies. To see whether according to the ontic reading (96) is true, we have to calculate  $Intervene_M(w, \neg R_1) \models P$ , whereby  $w$  maps  $C$  on 1,  $O$  on 1,  $R_1$  on 1,  $R_2$  on 1, and  $P$  on 1. To keep the presentation at a reasonable size, figure 5.12 lists only those worlds of  $U$  where the court orders the execution and the officer gives the signal. By now it would be clear that worlds where the causal history of the antecedent changes are very far away according to the order and play no role for the interpretation. As the figure shows,  $Intervene_M(w, \neg R_1) = \{w_4\}$ . At this world rifleman 1 does not shoot, but rifleman 2 does and the prisoner dies. Again, we correctly predict that (96) is true according to its ontic reading.



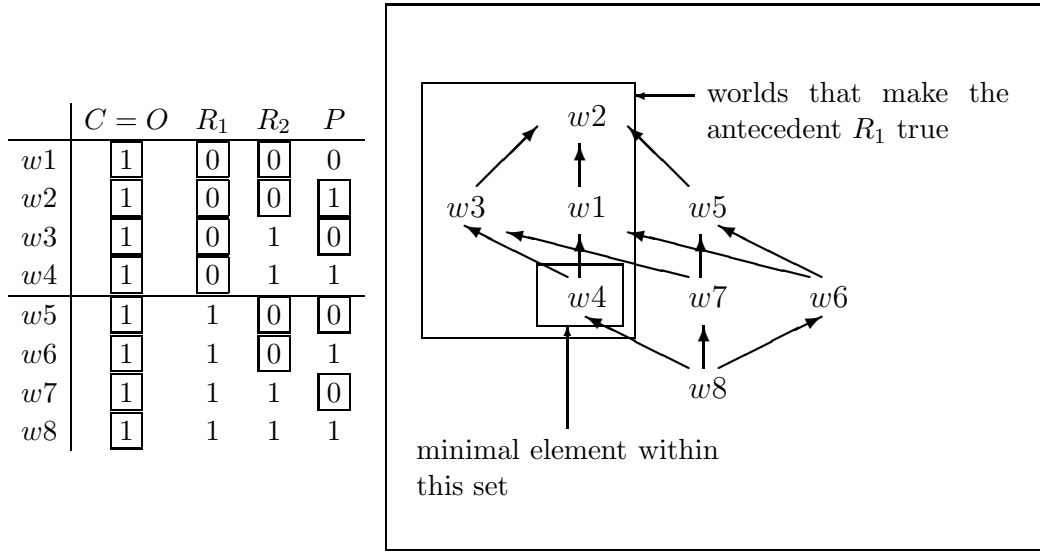


Figure 5.12: Minimal worlds for the shooting squad example

We see that the approach introduced here can account for a number of examples that are problematic for other approaches to the meaning of *would have* conditionals. But Pearl (2000), taken as a description of the ontic reading of *would have* conditionals, can account for these examples as well. What exactly is the relation between these two approaches? In spirit both proposals are very similar. The development of the present approach was strongly inspired by the ideas of Pearl (2000). But there are also a number of differences in the way these ideas are spelled out here. This should also be expected, given the number of problems we noticed for Pearl's approach. A first difference between Pearl's theory and the present proposal is that Pearl claims that his causal interpretation of conditionals applies to *would have* conditionals in general, i.e. there are no non-causal *would have* conditionals. On the contrary, we take the ontic reading only to capture one of two possible interpretations of such sentences. In consequence, many examples that Pearl cannot account for, our ontic reading does not have to account for, because we can explain them using the epistemic reading of *would have* conditionals. A second very obvious difference is that in the present framework the ontic revision function *Intervene* works by selecting minimal models. Pearl, on the contrary, provides a constructive description of the output of revision. However, he also suggests that we see this output as result of some process of selecting maximally similar models. So, one might say that the formalization of the revision function we propose is much closer to Pearl's intuitions about the way intervention works than his own proposal. Another important difference is that the objects affected by the revision operation *Intervene* are here, not – as in Pearl's theory – the causal structure but interpretation functions/possible worlds. Therefore, in contrast to Pearl's (2000) approach, the proposal presented here does not predict in a strict sense that to make the antecedent of ontic conditionals true laws

have to be given up. Causal laws might be violated in a particular world. But this does not affect what counts as a causal law according to the overall model. This way we can overcome some problems that we discussed for Pearl's approach in section 5.5.4. For instance, there we observed that manipulating the causal laws directly can lead to problems in more extensive frameworks where different instantiations of one and the same law can be distinguished. This problem is solved now. We have to say *fortunately*, because in the time-sensitive framework of the next chapter indeed the same causal law can apply at different times in one and the same possible world. We also solve two other problems of Pearl's approach this way. First, we are no longer restricted to conditionals whose antecedent contains only endogenous variables. Second, we are also no longer bound to antecedents that are conjunctions of literals. The theory developed here can deal with antecedents that may be arbitrary elements of  $\mathcal{L}^0$ . In section 5.5.4 we discussed another problem for Pearl's approach. His theory has been criticized because it assumed causal laws to be deterministic: given any valuation of the causes, the value of the effect is always determined. We saw that there are examples where the underlying laws do not appear to be deterministic. The theory developed here is not bound to deterministic causal laws and can handle these examples. This has been illustrated with the Tichy example that is treated correctly by the present approach. A less obvious difference is that the approach presented here predicts more ontic conditionals to be valid, because it respects information from analytical/logical laws. However, Pearl's approach can be easily extended in such a way that he can deal with this kind of information as well. Finally, we have to admit that one of the problems noticed for Pearl's approach is still unsolved by the present proposal. We still cannot handle correlations that cannot be analyzed as analytical/logical laws, such as the relation between the barometer and the probability of storm that was relevant for the examples (78a) and (78b) of section 5.5.4. This issue has to be left for future research.

An important prediction of the description of the ontic reading proposed here is that backward reasoning with causal laws is not possible for ontic conditionals. This is a consequence of two predictions of the approach. First, we allow for worlds where causal laws are broken. Second, we predict that worlds where causal dependencies are broken that immediately lead to variables occurring in the antecedent will always be more similar than worlds where the interpretation of causal ancestors of proposition letters occurring in the antecedent is changed. But two remarks on this general statement are in order. First, backtracking based on analytical truths is predicted to be possible. Hence, the approach predicts our first generalization for backtracking given in section 5.3.1. Let us illustrate this point. The *would have* conditional (97), repeated from section 5.3.1, is an example of a backtracking conditional that is generally found acceptable, even without adding an additional modal *have to* to the consequent. We propose that this

conditional is true in the ontic reading<sup>58</sup>, because it is based on an analytical law. This law tells us what the relation is between the age of a person and his year of birth. Analytical laws have to hold in all worlds. Hence, no matter which worlds *Intervene* selects to make the antecedent true, the consequent will be true at them as well.

- (97) If Clarissa were 30 now, she would have been born in 1966. (Frank 1997: 297)

Second, we do not predict that an ontic conditional  $\psi \gg \phi$  where all proposition letters in  $\psi$  causally depend on proposition letters in  $\phi$  and no analytical/logical laws support backward reasoning cannot be true. The approach predicts instead that in this case  $\psi \gg \phi$  is true if and only if  $\phi$  is true. The reason is that in such a situation the revision leaves the interpretation of atomic propositions occurring in  $\phi$  untouched. This prediction of the present approach implements Lewis' (1979) informal proposal for the truth conditions of backtracking counterfactuals (Lewis, 1979: 458) and also accounts for the second subclause of the Harper paradigm for the interpretation of *would have* conditionals. In fact, the present approach can be seen as a formalization of the spirit behind the Harper paradigm, though it deviates from it in the details.

#### 5.6.4 Discussion

In the previous section a new proposal for the meaning of *would have* conditionals was introduced. This approach distinguishes between two interpretations for *would have* conditionals. First, there is an epistemic reading. According to this reading a *would have* conditional is true if you would believe the consequent in case you learned that the antecedent is true. Hence, this reading is based on belief revision and follows the Ramsey receipt. The second reading is the ontic reading. The ontic reading is about what would be the case if the world itself were different. It reasons about the consequences of a hypothetical modification of the facts or reality rather than the hypothetical modifications of your beliefs about the facts.

This ambiguity of *would have* conditionals is realized in the presented framework by introducing two different conditional operators  $>$  and  $\gg$  into the formal language. For these operators independent interpretation rules are given. Conceptually, these interpretation rules are very similar. Both apply a revision function to the evaluation world and the antecedent to calculate the worlds at which the consequent has to be true to make the conditional true. In both cases revision is modelled by selecting minimal elements with respect to some order. Furthermore, both readings are instantiations of a particular subclass of similarity approaches: premise semantics. Premise semantics distinguishes a characteristic set of facts

---

<sup>58</sup>It is true in the most obvious epistemic reading as well.

of a world (or a belief state) on basis of which the order for the revision function is defined. Traditionally, one set of such premises is distinguished and the order is defined as maximizing the overlap of the premises with those of the evaluation world (or the relevant belief state). More sophisticated approaches like Veltman (2005) distinguish different premise sets that might influence the order in different ways. The way we formalized the epistemic and the ontic reading of conditionals follows the lines of Veltman (2005) who (with Goodman 1955) distinguishes three different set of facts of a world (a belief state) that are all of different relevance for the order: (i) the general laws that are taken to hold, (ii) the accidental facts of a world/a belief state from which together with the laws all other facts can be derived, and (iii) the facts that can be derived this way.

So much about what the interpretation rule of the epistemic reading and the interpretation rule of the ontic reading have in common. But what are the differences? First, the models that are minimized in both cases are of a different type. The ontic reading is about hypothetical changes of the world, while the epistemic reading reasons about hypothetical belief states. Hence, in case of the ontic reading worlds are compared, while for the epistemic reading the order applies to belief states. Second, there are differences in how exactly the three premise sets are defined. The set of laws contains all laws in case of the epistemic reading, while for the ontic reading it is restricted to analytical/logical laws. Causal or natural laws may be disobeyed by the worlds compared. The two readings also make use of different sets of basis facts. In case of the epistemic reading the basis is the set of facts for which the agent of the belief state has (independent) external evidence. In case of the ontic reading it is the set of initial conditions of a world. Finally, the basic facts and the third set of premises, the facts that are derivable from the basis and the laws, count in different ways for the similarity relation of both readings. While the epistemic reading follows Veltman (2005) in that it selects for maximal overlap of the bases and takes the third set of premises to be of no relevance for the order, the ontic reading maximizes overlap of the bases, and, additionally, minimizes the differences with the basis of the evaluation world. Furthermore, non-basis facts are also taken to be relevant for the order by the ontic reading.

A natural question this approach raises is how the two readings proposed for the meaning of *would have* conditionals interact with each other. It is important to realize that for many uses both readings of *would have* conditionals are in fact predicted to be identical. This is the case if the conditional and the context meet the following conditions:

- (i) the antecedent does not causally depend on the consequent, and
- (ii) the basis for the epistemic reading contains only atomic sentences that are

causally independent of the antecedent and vice versa.<sup>59</sup>

If these two conditions are met, then the two readings are predicted to be the same. If, however, one of these points is violated, then the proposed readings for a *would have* conditional differ. But this prediction seems to be in accordance with intuitions. Examples where the first condition is broken and indeed two different readings can be distinguished are all cases of explicit causal backtracking. Examples for a violation of condition (ii) are the Duchess example, the Hamburger example, and the Kennedy example. In all three cases some facts in the basis of the epistemic reading are not atomic sentences. Also the King Ludwig example belongs to this group. In this case some of the facts in the epistemic basis are not causally independent of the antecedent.

However, not in every context are both readings available. As everywhere else in natural language, in this case the context, particularly what counts as relevant information, can disambiguate a *would have* conditional. For instance, in the natural context where we place the King Ludwig example: spoken by some observer of the castle from a distance, it is not relevant what would have *happened* when the flag were flown, but what would we have *learned* if we had seen the flag up. Hence, the epistemic reading of the conditional is the most appropriate for this context. On the other hand, in the shooting squad scenario the relevant issue is not what we would have *inferred* if somebody had told us that rifleman one did not shoot, but what had *happened* if he didn't shoot. Thus, in this case the ontic reading is pragmatically dominant.

Still, there is something that has to be explained. There seems to be some imbalance between the ontic and the epistemic reading. For one thing, out of context the ontic reading appears to be the dominant reading. Second, when asked to judge based on examples of *would have* conditionals for which the proposal made here predicts different truth conditions for ontic and epistemic reading, people tend to argue about whether the predicted epistemic reading exists at all. Some people consequently deny its existence, other people are not sure about its existence. These observations suggest that there is something deficient or weak about the epistemic reading. How can this be explained? The epistemic reading, because it is based on belief revision, makes a statement about the epistemic state in which the conditional is evaluated. In particular, it makes a claim about the facts for which this belief state has external evidence. That means that discourse participants can only agree on an epistemic conditional if they share the same evidential history of their belief states. Only in very special contexts will this be warranted: for instance, if how the relevant information is given has been discussed explicitly in the context. Furthermore, the epistemic reading is not stable. That means that if the set *B* of facts for which an agent has external evidence

---

<sup>59</sup>One has to allow for the possibility that the antecedent or its negation itself to be part of the premise set. But I think that this does not stand in conflict with the formulation of this condition, because causal dependence is not a reflexive relation.

increases, then conditionals that have been true with respect to this belief state may become false. More formally, the following monotonicity condition does *not* hold: if  $K(w) = \langle B, U \rangle$  and  $K(w') = \langle B', U \rangle$  with  $B \subset B'$ , then  $w \models \psi > \phi$  implies  $w' \models \psi > \phi$ .<sup>60</sup> These observations may explain why the epistemic reading is only available on special occasions and strongly context dependent.

As explained at the beginning of this chapter, it has often been proposed that the past, in particular the past of the antecedent plays a special role for similarity. The intuition behind this claim is that “... in reasoning from a counterfactual supposition about any time, we ordinarily assume that facts about earlier times are counterfactually independent of the supposition and may freely be used as auxiliary premises.” (Lewis: 1979: 456). We have discussed some ideas for how to specify the similarity relation in a way that gives the past prominence for similarity – by demanding, for instance, that except for some minimal transition period most similar worlds have to be identical for the past of the antecedent, or that they have to be identical up to the decision time of the antecedent. We will discuss another approach along these lines in the next chapter. As we have concluded here, so far we see no simple way to give an exact implementation of this idea that does not run into obvious empirical problems. Our approach can explain the basic intuition that counterfactual reasoning (normally) does not change the past. It is a consequence of the important role causality plays for the ontic reading. The ontic reading reasons from causes to their effects. Therefore, for the ontic reading it is indeed true that earlier times are counterfactually independent of the supposition.

There is one aspect of the present approach that can be expected to raise questions by some readers. Together with Pearl (2000), Veltman (2005), and many other approaches towards the meaning of *would have* conditionals, we describe the truth conditions of these sentences as referring to a set of (contextually salient) laws. However, we do not provide a way to calculate the relevant laws for a concrete occurrence of a *would have* conditional. We simply stipulate the relevant law structures when discussing examples and hope that our intuitions on this point are shared by those of the reader. Of course, it would be better, if we could provide a theory of what makes some generalizations laws that we could build on with our approach to the meaning of *would have* conditionals. But to develop such a theory is a topic different from those addressed in this thesis and would lead us far beyond the scope of the present work.

Our law-based approach may also attract questions from people who are convinced – following Lewis – that counterfactuality is conceptually prior to causality. Our theory for the meaning of *would have* conditionals might be suspected to lead to circularity when applied to a theory of causation along the lines of Lewis pro-

---

<sup>60</sup>For an example take  $B = \{A\}$ ,  $B' = \{A, B\}$ , and the conditional  $(\neg A \vee \neg B) > \neg B$ .

positional. But also this kind of criticism misses the purpose of the thesis. The aim of the research presented here is to account for the meaning of English conditional sentences. We have shown that a theory that bases this meaning on a given set of (causal) laws solves many issues about the truth conditions of *would have* conditionals that are problematic for various other approaches. The fact that this may not fit in some philosophical theory about the nature of causation is not relevant for the linguistic objective of the thesis.<sup>61</sup>

Above we have argued for two potential lines of criticism against the present approach that they do not apply. Let us finally point out an aspect of the proposed theory that we think may turn out problematic under closer consideration. Important steps in our description of the ontic reading of *would have* conditionals are, first, to make a distinction between analytical/logical laws on the one hand and causal laws on the other, and, second, to allow reasoning based on causal laws only to go in one direction: from the cause to the effect. Although the distinction of a group of laws that come with a direction appears to be crucial for a correct description of the ontic reading of *would have* conditionals, it may turn out that the notion of *causal laws* does not provide a complete characterization of this group. For some examples causality does not seem to be the right classification of the underlying law. A paraphrase of the form *A is a reason for B* appears sometimes much more proper than *A is a cause of B*. This line of thought has to be continued in future research.

## 5.7 Summary

The objective of the sixth chapter was to come up with a convincing description of the meaning of *would have* conditionals. This description abstracted away from two aspects of these sentences. First, the compositional structure of the sentences was to be ignored, except for the distinction between antecedent and consequent, where both are treated as ordinary, in particular unmodalized statements about the world. Second, the description ignored (to a large extent) temporal aspects of antecedent and consequent.

---

<sup>61</sup>It does not lie within the scope of this work to get involved in the philosophical discussion about what should be taken to be primitive, causality of counterfactuality. But let us add that at least it is hard to defend that counterfactuality has *cognitive* priority to causality; in other words, that we compute causal relationships based on our ability to compute counterfactual reasoning. Various observations seem to support the idea that causality is in some sense innate. There is evidence that a disposition to distinguish between certain causal and noncausal sequences is widely shared among humans and many nonhuman animals, emerges early in development, and in some cases is remarkably fast and efficient. Human children appear to be able to recognize causal dependencies at a very early age (see, for instance, Leslie & Keeble 1987). On the contrary, counterfactual reasoning has been shown to be very hard for young children (Riggs et al. 1998 and Peterson & Riggs 1999) and to be acquired much later than the understanding of causal dependencies.

Before the new approach was introduced we first discussed various other proposals made to motivate the chosen account. We started with the similarity approach to the meaning of counterfactuals brought forward by Stalnaker (1968) and Lewis (1973) (section 5.2). They propose that a *would have* conditional is true if on those models for the antecedent that are most similar to the evaluation world the consequent is true as well. The central problem of this line of approach is the vague description of the similarity relation. It has been argued in the literature that if the similarity approach is correct, similarity is not so semantically underspecified as assumed by Stalnaker and Lewis. There are general restrictions on what makes a world being closer to the evaluation world than some other world.

One restriction that has been proposed particularly often is that the past plays a dominant role for similarity. It has been suggested at various places in the literature that similarity is similarity of the past, or even identity of the past – where different authors define past in different ways. We have argued that two central arguments brought forward to support this claim, (i) the issue of backtracking *would have* conditionals (section 5.3.1) and (ii) the future similarity objection (section 5.3.2), are not convincing as arguments to this point. Furthermore, we have argued that approaches that take similarity to be reducible to similarity or identity of the past have to face empirical problems.

We then discussed a particular subtype of similarity approaches: premise semantics. Premise semantics distinguishes a characteristic set of facts of a world (or a belief state) on the basis of which the order for the revision function is defined. We have focused on one specific approach along the lines of premise semantics: Veltman (2005). This approach distinguishes, with Goodman (1955), two premise sets that are relevant for the order: (i) the general laws that are taken to hold, and (ii) the accidental facts of a world/a belief state from which, together with the laws, all other facts can be derived. The second set is called the *it basis* of a world. The proposed similarity relation then selects those worlds making the antecedent of a conditional true where (i) all general laws holds and (ii) a maximal set of basis facts holds. A strong advantage of this approach is that it is able to make precise predictions for concrete examples. This is something many other similarity approaches miss. Furthermore, these predictions turn out to be correct in many cases. However, we have also seen that there are a number of examples that the approach cannot account for.

We then turned to an approach to the meaning of *would have* conditionals that comes from a totally different background (section 5.5). Pearl (2000) claims that counterfactuals should be interpreted as executing hypothetical surgeries on the causal network governing the evaluation world. He proposes the following evaluation strategy for the evaluation of *would have* conditionals. First, one cuts all causal dependencies that connect the antecedent to facts causally responsible for its falsity. Second, one simply stipulates the truth of the antecedent as a causal law. Finally, one checks whether from this new causal network the truth of the



consequent can be derived. It turns out that with this theory one can predict correct truth conditions for many examples problematic for the other accounts discussed before. One question this approach raises is what the exact relation is with the similarity framework. Pearl (2000) shows that there is a close connection with the logic of Lewis' (1973) counterfactual, but he does not provide a definition of his approach in terms of similarity. On these grounds it is difficult to relate this proposal to the approaches discussed earlier. The central problem of Pearl's approach is that there is a huge class of examples he cannot account for, namely those *would have* conditionals whose truth is not based on causal dependence between antecedent and consequent.

At this point a new approach to the meaning of *would have* conditionals was introduced (section 5.6). It was proposed that there exist two interpretations for *would have* conditionals. First, there is the epistemic reading. According to the epistemic reading a *would have* conditional is true if you believed the consequent in case you learned that the antecedent is true. Hence, this reading is based on belief revision and follows the Ramsey test. The second reading is the ontic reading. The ontic reading is about what would be the case if the world itself were different. It reasons about the consequences of a hypothetical experiment on reality rather than the hypothetical modifications of your beliefs about the facts. Both readings follow premise semantics and Goodman's receipt for the meaning of *would have* conditionals. A set of laws and a basis of accidental facts are distinguished. The epistemic reading applies premise semantics to belief states. Thus, the basis is a characteristic set of facts for a belief state. We propose that it is the set of facts for which the agent of the belief state has independent external evidence. According to the epistemic reading a *would have* conditional with antecedent  $\psi$  and consequent  $\phi$  is true, if the consequent holds on the belief state you obtain by keeping all laws and a maximal subset of the basis of the evaluation belief state. The ontic reading applies to worlds instead of belief states. Thus, now the basis is a characteristic set of facts about a world. This set is described as the facts of the evaluation world from which all other facts can be derived from general laws, whereby reasoning on casual laws has to go from cause to effect. According to the ontic reading a *would have* conditional with antecedent  $\psi$  and consequent  $\phi$  is true if the consequent holds on those worlds that (i) obey all analytical/logical laws, (ii) have a maximally similar basis to the evaluation world, and (iii) are maximally similar with respect to the facts derived from the laws and the basis. We have seen that this approach allows us to deal with those examples that are problematic for the other proposals discussed. This gives us hope that we have made an important step towards a correct description of the notion of similarity involved in the interpretation of *would have* conditionals.

## Chapter 6

---

# Tense in English conditionals

### 6.1 Introduction

The primary aim of this chapter is to account for the interpretation of the tenses and – as far as its temporal properties are concerned – the perfect in English conditional sentences. In particular, we want to address the question whether we can account for the interpretation of the tenses and the perfect in these constructions in a compositional way. In the second place, the approach should be faithful to the results of the last chapter. That means that the compositional theory for conditionals that is developed here should produce the same semantics for *would have* conditionals as proposed in Chapter 5. We will not be able to directly transfer the proposal made there into the present framework. The introduction of time into the model and the more complex formal language will make some small amendments necessary. But these amendments will not affect the predictions made for the meaning of *would have* conditionals we discussed in the previous chapter.

The interpretation of the tenses in conditionals is a challenging topic that has fascinated and puzzled many philosophers and linguists in the past. Conditionals demonstrate temporal properties that stand in conflict with what you would expect given the temporal and aspectual operators occurring in them. It is not our aim to give a complete survey of the temporal properties of conditionals or the literature on this issue. We will focus on two surprising temporal features of conditional semantics and try to account for them:

- (i) *the puzzle of the missing interpretation*, and
- (ii) *the puzzle of the shifted temporal perspective*.

The first puzzle may be the best known and discussed puzzle concerning the interpretation of the tenses and the perfect in conditionals. The observation is that in subjunctive conditionals the simple past and perfect markings in antecedent

and consequent appear not to be interpreted. For illustration, in the antecedent of the indicative conditional (98a) the finite verb is marked by the simple past. As we would expect given standard theories about the meaning of the simple past in English, the antecedent refers to a situation in which Peter left at some time in the past. In example (98b) we have a *would* conditional, again with an antecedent whose finite verb is marked by the simple past. But this sentence is semantically anomalous. The antecedent of (98b) cannot be about some past situation, but has to be about the present or the future. This is incompatible with the restrictions on the evaluation time of the antecedent introduced by the temporal adverbial *yesterday*.

(98) a. If Peter left yesterday, he will be in Frankfurt this evening.

b. \*If Peter left yesterday, he would be in Frankfurt this evening.

Two ways of approaching this problem can be distinguished in the literature. Firstly, it has been proposed that, even though it does not look that way, the tense and aspect morphology in subjunctive conditionals carries in this context the same meaning as in simple sentences. Proposals along these lines all follow roughly the same idea: the past or the perfect do not shift backward the evaluation time of the antecedent or consequent, but the evaluation time of the conditional as a whole. The price paid for being able to stick to the standard meaning for the tense- and aspect morphology in subjunctive conditionals is, thus, a logical form that does not follow the surface structure of the sentences. On the surface, there is no past on the top of conditional sentences. As we will see, approaches along these lines often can only account for parts of the puzzle of the missing interpretation. That means they can account for either the past tense or the perfect, and sometimes additionally only for the occurrence of the past tense or the perfect either in the antecedent or in the consequent. Furthermore, we will argue that the underlying idea of these proposals does not lead to a convincing description of the meaning of subjunctive conditionals. Therefore, we will dismiss this approach to the puzzle of the missing interpretation in general. Alternatively, it has often been claimed that the simple past or the perfect has a mood/modality meaning in subjunctive conditionals. The criticism many proposals along this line have to face is that they describe the meaning of the aspect and tense morphology in conditionals only in very vague terms. As a consequence, they make rather diffuse predictions for the semantics of these sentences and other constructions containing the same tense and aspect markings. In section 6.4 we will provide a compositional semantics for subjunctive conditionals that adopts the past-as-modal approach to the puzzle of the missing interpretation, but makes very specific claims about the meaning of the simple past and the perfect and the way they contribute their meaning to the interpretation of conditionals.

The second temporal property of conditionals we want to account for concerns in the first instance the interpretation of the tenses in indicative conditionals. It

is quite generally accepted that the meaning of the English tenses has a deictic element. They locate the evaluation time of the sentence they modify relative to the utterance time: the simple present locates the evaluation time at the utterance time (or in its future), the past locates this time before the utterance time. However, this appears to be falsified by indicative conditionals. A past tensed consequent in such a conditional can sometimes be evaluated in the future of the utterance time (see (99a) and (99b)).

- (99) a. If Peter comes out smiling, the interview went well.  
       b. If the package arrives tomorrow morning, it was posted this evening.

Something similar holds of the simple past occurring in relative clauses of modal sentences.

- (100) a. I will eat a fish that was alive.  
        b. I might marry a man that was in prison.

The puzzle of the shifted temporal perspective is much less discussed in the literature than the puzzle of the missing interpretation, but it is just as intriguing. One way to look at it is to see it as a variation of the first puzzle: the past tense is not interpreted how and where you would expect it to be. More in accordance with the observations is the view that the past tense is interpreted according to standard lines, but that the reference time of tenses is not obligatorily the utterance time. In conditional and modal contexts the reference time can be set to locations in the future of the utterance time. We will follow this second view and propose that the shifted temporal perspective in conditional and modal contexts is a direct consequence of the update conditions for the ontic reading of antecedents of conditionals and modals.

The chapter is structured as follows. In the next two sections both puzzles concerning the temporal behavior of conditionals are discussed in more details. We will introduce some approaches made to explain these puzzles and evaluate them. Afterwards a compositional approach to conditionals will be proposed. It will be shown that this approach can solve both puzzles. We conclude the chapter with a discussion of this new approach and summarize the findings.

## 6.2 The puzzle of the missing interpretation

### 6.2.1 The observations

Let us take a closer look at the puzzle of the missing interpretation. If you look at the form of *would* conditionals and pay particular attention to the syntactic

tense and aspect markings of these sentences, then you see that the finite verb in antecedent as well as consequent is marked for the simple past.<sup>1</sup> Similarly, *would have* conditionals show a simple past marking on the finite verb in antecedent and consequent followed by a syntactic perfect formed by *have* plus a past participle. A first outline of a syntactic structure for subjunctive conditionals respecting these observations, that also tries to stay as close as possible to surface structure, would appear as described in figure 6.1.<sup>2</sup>

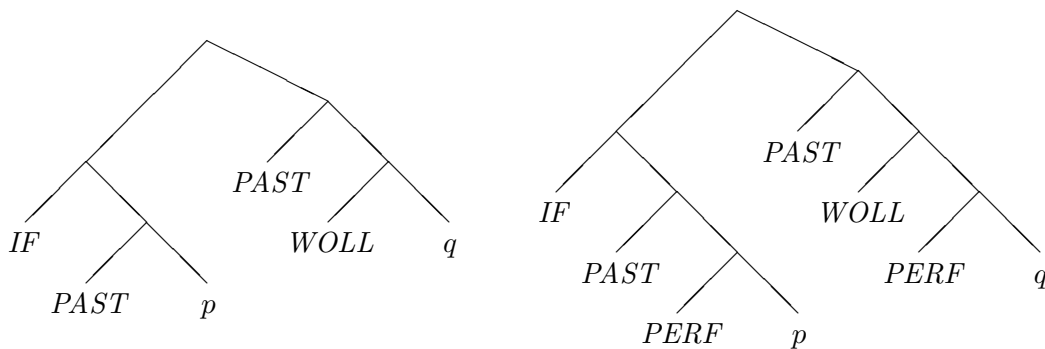


Figure 6.1: A simple syntactic analysis of *would* and *would have* conditionals

Let us consider what predictions for the temporal properties of subjunctive conditionals we would make if we combined this structure with standard approaches to the meaning of the simple past, the perfect and the modal *WOLL*. Standardly, the simple past is interpreted as localizing the evaluation time of the sentence it is attached to at some contextually given time point before the utterance time, while the perfect is interpreted as localizing the evaluation time of the phrase in its scope at some point in the past of the evaluation time of the perfect. We also need to say something about the temporal properties of the modal. It is often proposed that the evaluation time for the phrase in the scope of the modal is localized at or in the future of the evaluation time of the modal. Taking everything together, this means that under assumption of the structure described in figure 6.1, a conditional like (101) can be paraphrased as follows.<sup>3</sup>

<sup>1</sup>The position that *would* is syntactically the past form of *will* is held by many authors (Palmer 1986, Comrie 1985, Quirk et al. 1985). It is less clear, whether *might* is the past form of *may*, because *might*, in contrast to *would*, cannot occur in contexts where it means past time reference. However, in earlier stages of English, both modals were used frequently in past tense contexts. I adopt the position that at least on the level of form both modals are marked for the simple past.

<sup>2</sup>The structure is decomposed to a level where the place at which the past tense and the perfect make their contribution in the logical form becomes clear, but not beyond this level. The remaining sentence radicals are denoted by small letters.

<sup>3</sup>Depending on which theory for the interpretation of the simple past and the perfect is adopted, relations between the temporal variables introduced may be predicted.

(101) If Peter took the plane, he would be in Frankfurt this evening.

*If at some contextual given point  $t$  in the past Peter took the plane, it follows that at some contextually given point  $t'$  in the past it would be the case that at some time  $t'' \geq t'$  this evening Peter is in Frankfurt.*

This cannot lead to a correct description of the meaning of this conditional. The antecedent of a *would* conditional can never refer to the past, but always refers to the present or the future. That the interpretation of the consequent does not follow this description is less obvious. The predictions made for the evaluation time for the modal are difficult to confirm or falsify based on intuitions about the meaning of (101). But it is intuitively clear that the evaluation time for the phrase in scope of the modal cannot lie in the past. The approach sketched above, however, would predict that it should be possible to evaluate this phrase in the past – the only constraint so far is that this time does not lie before the past evaluation time of the modal. Thus, we see that the predictions of this standard approach to the meaning of the past tense in conditionals do not match the observed temporal properties. It rather looks as if, semantically, there is no past tense active in the antecedent and consequent of these conditionals.

Let us turn to *would have* conditionals now. Also in this case the meaning predicted by combining the surface structure given in figure 6.1 with the standard approach to the past tense and the perfect is not correct. According to such an approach a sentence like (102) would mean something like the following.

(102) If Peter had taken the plane, he would have been in Frankfurt this evening.

*If at some point  $t'_1$  in the past of some contextually given past time  $t_1$  Peter had taken the plane, it follows that at some contextually given point  $t_2$  in the past it would be the case that at some point  $t'_2 \geq t_2$  the perfect statement in scope of the modal is true, i.e. there is some time  $t''_2 < t'_2$  this evening where Peter is in Frankfurt.*

Let us focus on the antecedent, because there the problem is most transparent. While the antecedent of a *would have* conditional is often evaluated in the past, it does not refer to the past of some past time, as do standard past perfect constructions. Furthermore, it has been noticed by many authors that *would have* conditionals can also be evaluated in the future or at the present (see, for instance, Jespersen 1924, Dudman 1984, Leirbukt 1991). The following examples illustrate this possibility. The first one is due to Leirbukt (1991). He mentions a daily soap as his source. Together with the second sentence it exemplifies the possibility of *would have* conditionals to refer to the future. The last sentence shows that reference to the present is possible as well.

(103) a. I'm glad that you called. In a quarter of an hour I would have been gone.

- b. If you had called in a quarter of an hour, I would have been gone.
- c. Unfortunately, Peter left us the other day. But if he had been here now, he would have been terribly glad to see you.

Also for *would have* conditionals it looks as if the past tense is not interpreted as such. One even gets the impression that the same is true for the perfect as well. This misfit between what standard approaches to the simple past and the perfect predict and the actual temporal properties we observe for subjunctive conditionals – where their semantics seems to have no effect – is what we call the puzzle of the missing interpretation.

Before one can start to look for an explanation of this puzzle, it is important to realize that the observed mispredictions made are not only a result of the adopted meanings for the past and the perfect, but also the logical structure assumed for subjunctive conditionals. This suggests that we distinguish two ways to approach the puzzle of the missing interpretation. First, one could claim that the proposed syntactic structure that governs compositional semantic is false. Such a position may try to maintain the standard meaning for the past and the perfect, proposing that they contribute their meanings not in the way and at the place that would follow from the trees in figure 6.1 on page 162. We will call approaches that follow this line *past-as-past* approaches. A different option is to say that surface structure describes correctly the place where the tenses and the perfect contribute their meanings, but the meanings assumed by standard semantics for the past and the perfect are not correct – at least in the context of subjunctive conditionals. This is the strategy that most authors discussing the puzzle of the missing interpretation strategy follow. Approaches along this line will be called here *past-as-modal* accounts, because they often propose a modal meaning for the simple past (and sometimes also the perfect) in conditional contexts. In the following, we will discuss a number of proposals following either the *past-as-past* strategy or the *past-as-modal* strategy. We will discuss their respective potential, but also the problems they come with and thereby set the basis for the explanation of the puzzle of the missing interpretation that will be proposed in section 6.4.

### 6.2.2 Past-as-past approaches

Past as past approaches are conceptually very attractive. They have the potential to maintain the standard meanings for the simple past and the perfect. This is interesting, because changing this meaning, particularly introducing a lexical ambiguity would complicate the lexicon. Additionally, one would like to maintain the standard meanings, because in many sentences they make the correct predictions. We have said above, that the price to be paid for a conservative lexical semantics is to give up the syntactic analysis sketched in figure 6.1 on page 162. But why should we adopt this analysis? One has to admit that this analysis was

rather naive and might very well be wrong. Furthermore, what does commit us to a logical form that mirrors surface structure? Issues like quantifier raising have forced us to become used to the idea that this does not always have to be the case.

Despite its attractiveness, there are only few semanticists that have tried to give substance to the past-as-past approach. One reason is that the alternative past-as-modal approach has a lot of intuitive appeal. But this is certainly also the consequence of the difficulties one has to face when one tries to work out a past-as-past proposal. It is easy to come up with suggestions for different structures for conditional sentences, but much more difficult to find one that explains the puzzle of the missing interpretation. One of the few works following the past-as-past hypothesis is Tedeschi (1981). The essential idea behind his approach is that the simple past in subjunctive conditionals does not apply to the eventualities described in antecedent and consequent, but to the conditional as a whole. Hence, the semantical structure of such a conditional looks rather as follows.

Tedeschi's interpretation rule for subjunctive conditionals

A subjunctive conditional with the tenseless propositions  $p$  in the antecedent and  $q$  in the consequent is assigned the following logical form:

$$P(p \succ F(q)),$$

where  $P$  and  $F$  are logical operators shifting the evaluation time backward ( $P$ ) and forward ( $F$ ) respectively, and  $\succ$  a conditional connective, whose meaning still has to be defined.

From a compositional perspective, this proposal is not very convincing. It is not clear why the past operator is in the position superordinating the conditional. Furthermore, the approach is bound to an analysis of the simple past as a sentential operator. Many students of tense in English have argued that this is not the way English tenses work (see, among others, Kamp & Reyle 1993). However, the underlying idea, that subjunctive conditionals are conditionals evaluated in the past has a long tradition in the literature of conditionals. It is actually the leading idea of all past-as-past approaches. It is also very often taken as a basis for the semantic meaning of *would have* conditionals by authors that want to derive the counterfactuality of these conditionals as conversational implicature (see Condoravdi 2002, Ippolito 2003, and many others). But also philosophers have found it very attractive, for instance, to describe the difference between indicative and subjunctive conditionals (see Adams 1975, 1976, and Skyrms 1980, 1981, 1984, 1994).

Before we discuss more past-as-past approaches in detail, let us first clarify this common idea for the meaning of conditionals. The basic claim is that the class of conditionals can be split into those evaluated with respect to possibilities



admissible at the utterance time<sup>4</sup> (the indicative conditionals, for some authors also the *would* conditionals), and conditionals evaluated with respect to sets of possibilities accessible at some past time point (*would have* conditionals, for some authors also *would* conditionals). Let us call the first group *present conditionals* and the second group *past conditionals*. Two sets of possibilities are generally considered relevant for the evaluation of conditionals.<sup>5</sup> In one case the conditional is read epistemically. In this case, the possibilities are the possible worlds consistent with what some agent believes/knows at some time-point. According to the other reading, the conditional makes reference to the ontic (metaphysic) alternatives. Ontic alternatives are also represented by a set of possible worlds. But this time these worlds do not represent what is known or believed by some agent, but what is settled about some world. Intuitively, a fact is settled, if it is no longer open to manipulation, it cannot be changed. A central claim or observation about settledness is that it depends on time: while the past and the present of a world are settled, the future is – to some extent – still open. Kamp (1978) illustrates this difference with two games, GOF<sup>6</sup> and GOP<sup>7</sup>. GOF works as follows. It is played by two players A and D. A makes some claim about the immediate future and D has to respond by saying whether the claim is correct or not. It is easy to see that player A has a winning strategy in this game: the player just makes some claim about some fact concerning the future that is under his control. For instance, that he will scratch his nose in a minute. Then D has no chance to get the answer right. The rules of GOP are similar to those of GOF. The only difference is that this time A has to make a claim about the immediate past. Now, it is not that easy for A to win. The reason is – or that is the claim – that A has no control about the past. The past is already settled.

The standard formalization of the notion of settledness or the ontic alternatives follows the branching futures approach introduced by Kamp (1978) and Thomason (1985). According to this formalization, time is a linear structure and worlds are complete histories, interpretation functions defined for the whole time line. Let  $T$  be the set of times and  $M$  be the set of complete histories over  $T$ . To model settledness, an accessibility relation  $\cong$  between worlds is introduced that at a certain time relates all those worlds that share the same history up to this time point. Hence,  $\cong$  is a 3-place relation on  $T \times W \times W$ , such that (i) for all  $t$ ,  $\cong_t$  is an equivalence relation, and (ii) for any  $w_1, w_2 \in W$  and  $t_1, t_2 \in T$ , if  $w_1 \cong_{t_2} w_2$  and  $t_1 < t_2$ , then  $w_1 \cong_{t_1} w_2$ . These conditions warrant that the set of worlds  $w'$  that stand in the relation  $\cong_t$  to a world  $w$  decreases over time. At each  $t$  the relation  $\cong_t$  splits the set of all worlds into equivalence classes that become smaller and smaller over time. Finally, one demands that if  $w \cong_t w'$  and  $P$  is a

---

<sup>4</sup>Adams (1975, 1976) suggests that these conditionals have to be evaluated with respect to a present epistemic probability distribution.

<sup>5</sup>Kratzer (1979, 1981) proposes many more.

<sup>6</sup>*Game Of the Future*

<sup>7</sup>*Game Of the Past*

atomic proposition letter, then  $\forall t' < t : t' \in w(P) \Leftrightarrow t' \in w'(P)$ . This condition assures that the worlds standing in the relation  $\cong_t$  do in fact interpret the past up to  $t$  identically.

Based on the relation  $\cong$  we can now give a rough formalization of the evaluation strategy for past conditionals described above. We will call this the *back-shift interpretation rule* of past conditionals.

The back-shift interpretation rule for past conditionals

A past conditional with antecedent  $A$  and consequent  $C$  is true in  $w_0$  at  $t_0$  if

$$\exists t < t_0 \forall w : (w_0 R w \ \& \ A(w)(t)) \Rightarrow C(w)(t),$$

where  $R$  is either an epistemic accessibility relation or the ontic accessibility relation  $\cong$ .

To illustrate this approach to conditionals with an example, consider again the Kennedy example we have discussed in Chapter 5.

(104) a. If Oswald didn't kill Kennedy, someone else did.

b. If Oswald hadn't killed Kennedy, someone else would have.

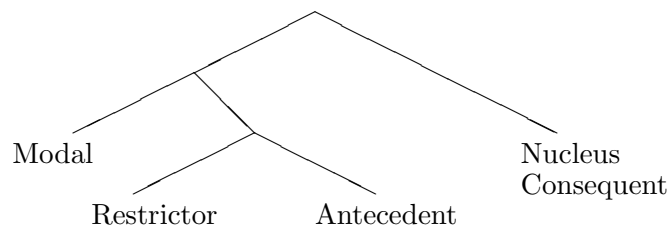
Most people agree with the first, indicative conditional, but deny the second, subjunctive conditional. An approach along the lines sketched above could now explain the difference in truth conditions as follows. Interpreters that share the intuitions just described believe that Kennedy is dead, but they may have doubts concerning whether Oswald was indeed the murderer. The first conditional is an epistemic present conditional, evaluated with respect to the epistemic alternatives of this interpreter at the utterance time. The interpreter does believe that Kennedy is dead. Hence, in all worlds consistent with his beliefs where Oswald did not kill Kennedy someone else has to be the murderer. The second conditional is an ontic past conditional, evaluated with respect to ontic alternatives accessible at some past time.<sup>8</sup> We go back in the actual world to some time when the antecedent was still not settled and look at all those ontic alternatives where the antecedent turns out to be true. If the interpreter believed in a conspiracy theory concerning the death of Kennedy, then there would be some past time – probably the time at which the conspiracy was set up – at which in all futures where Oswald does not kill Kennedy, somebody else does. Hence, the second conditional comes out as true. However, the type of normal interpreter that we refer to here does not believe in conspiracies. Hence, he would find no past time were all futures in which the antecedent turns out to be false, the consequent becomes true. Thus, the conditional is predicted to be false.

---

<sup>8</sup>Condoravdi (2002) even claims that this is the only reading possible.

Let us now come back to the puzzle of the missing interpretation. The aspect of this analysis that makes it so attractive to proponents of a past-as-past hypothesis is that it allows for at least some of the past markers in subjunctive conditionals to keep their temporal meaning. They are taken to express the back-shift of the evaluation time of the conditional. The question is whether we can provide some plausible compositional semantics for conditionals that produces the described back-shift interpretation rule for subjunctive conditionals. Furthermore, we have to see whether this approach can completely explain the puzzle of the missing interpretation.

One past-as-past approach that tries to answer these questions has been brought forward by Ippolito (2003). She builds on the theory for conditionals introduced by Kratzer (1979, 1981). Kratzer proposes that conditionals are modal statements. Modals, on the other hand, are according to Kratzer (1979, 1981) interpreted as quantifiers over possible worlds. They take two arguments denoting sets of possible worlds, a restrictor and a nucleus and then make a statement about the relation between these two sets. It is proposed that antecedents or if-clauses restrict the first argument, while the consequent describes the nucleus. According to this theory, a conditional has the logical structure described in figure 6.2. Kratzer proposes that the modal is not always explicitly present in a conditional, but is in standard conditionals the covert modal *Must*, interpreted as universal quantifier.<sup>9</sup>



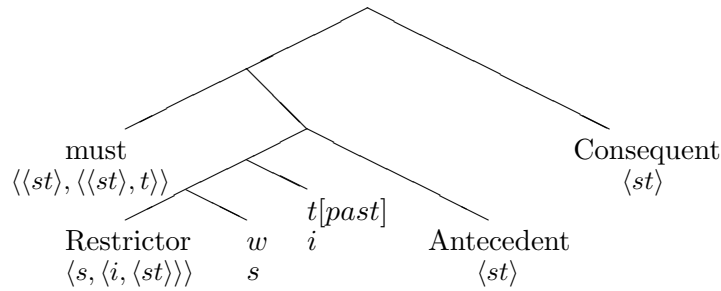
$$\forall w : (Restrictor(w) \& Antecedent(w)) \Rightarrow Consequent(w)$$

Figure 6.2: Kratzer's approach to conditionals

Ippolito (2003) only aims at accounting for *would have* conditionals that refer to the present or future. She calls these conditionals *mismatched past conditionals*. Remember, that in these conditionals both the simple past and the perfect seem to miss their temporal interpretation. Central for the approach of Ippolito (2003) is the claim that in a modal quantificational structure the past can be interpreted

<sup>9</sup>Kratzer seems to assume for all conditionals discussed here, also those with *will* and *would* in the consequent, that they contain a covert modal *Must*.

either in the restrictor or in the nucleus. She then proposes that in the case of mismatched past conditionals, it is interpreted in the restrictor. The restrictor is a time and world dependent accessibility relation. The simple past restricts the interpretation of the temporal variable of this accessibility relation to some time interval in the past.<sup>10</sup> According to her, this leaves the antecedent and the consequent as tenseless propositions. The structure Ippolito proposes for mismatched past conditionals is sketched in figure 6.3, together with what appears to be the meaning assigned to these sentences.<sup>11</sup>



A mismatched past conditional with antecedent  $A$  and consequent  $C$  is true in world  $w_0$  at time  $t_0$ , if the following holds:

$$\exists t < t_0 \forall w : [w \cong_t w_0 \ \& \ A(w) \Rightarrow C(w)]$$

Figure 6.3: Ippolito's approach to conditionals

The central problem of this approach is that it provides no explanation for why in a modal quantificational structure the past can be interpreted either in the restrictor or in the nucleus. Furthermore, Ippolito (2003) leaves unclear how *four* markers of past time reference – simple past and perfect in antecedent and simple past and perfect in consequent – are interpreted as one past operation, or, in other words, why interpreting one past feature on the restrictor turns the antecedent and the consequent into tenseless propositions. It seems reasonable that one instance of the past referring morphology may somehow apply to the restrictor. But if this is what she means, which one is the chosen one? And what happens to the others?

There exists a proposal very similar in spirit to Ippolito (2003), but worked out much more systematically and precisely: Condoravdi (2002). Actually, Condo-

<sup>10</sup>Ippolito (2003) proposes a standard interpretation for the simple past: it imposes restrictions on the interpretation of the temporal variables to which it applies, the variable has to be interpreted as some time before the utterance times. She also assumes that the perfect is in some contexts interpreted this way.

<sup>11</sup>Ippolito (2003) assumes that mismatched past conditionals always refer to the ontic modal base. The presentation simplifies her tense semantics, that is anaphorical/presuppositional.

ravdi is also not primarily interested in accounting for the puzzle of the missing interpretation. Instead, she wants to account for some aspects of the (non-root) meaning of the modals *may*, *might*, *will*, *would*, particularly when combined with the perfect, as in (105). But in combination with Kratzer's approach to conditionals, her theory can be extended to a proposal about the meaning of conditionals.

(105) Peter might have won the game.

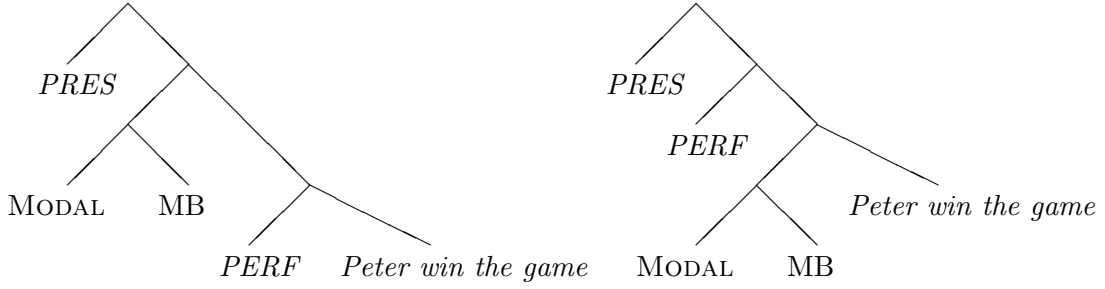
Because, in contrast to Kratzer, Condoravdi (2002) treats *will* and *would* as modals, many more conditionals than in the original approach of Kratzer now become explicitly modalized. The driving idea in Ippolito's proposal, that in modal structures some past feature can be interpreted either in the restrictor or in the nucleus of the modal, is also central for Condoravdi's approach. But now it is explained as a consequence of a structural ambiguity of the scoping relation between modal and perfect: the perfect can just as well scope over the modal as under it. Under the modal it applies to the phrase in scope of the modal, which describes the nucleus of the quantifying structure. Over the modal it shifts the evaluation time of the restrictor of the modal backward. This is all worked out in a detailed compositional semantics. Below, in figure 6.4, the two logical forms Condoravdi proposes for sentences like (105) are given, together with the meanings she proposes for the simple present, the perfect, and the modals.<sup>12</sup> The interpretation of the present tense and the perfect follows standard lines. The modals are analyzed as quantifiers over possible worlds. Their domain, the modal base *MB*, is contextually given. Condoravdi distinguishes two modal bases for the non-root readings of the modalities: an epistemic modal base and an ontic modal base.<sup>13</sup> They follow the description given above. The meanings she proposes for *might/may* and *would/will* combine a standard modal meaning with a temporal meaning: the modalities expand the evaluation time for the property of times in their scope forward. The last line in figure 6.4 describes the meaning Condoravdi (2002) predicts for (105) if the perfect scope over the modal.

In how far can this approach account for the puzzle of the missing interpretation? It is important to realize that – contrary to Ippolito's proposal – the structural ambiguity proposed applies only to the perfect and not to the past tense. Condoravdi can account for why it may sometimes look as if the perfect loses its interpretation in *would have* conditionals or modalities for the past. If the perfect scopes over the modal, then the evaluation time of the modal, more particularly the modal base, is shifted to the past. But because the modal expands

---

<sup>12</sup>We ignore here an important feature of her approach. Condoravdi distinguishes in her ontology sorted eventualities. Basic untensed sentences denote properties of eventualities. This enables her to account for some very intriguing facts about differences in interpretation between eventive and stative predicates. However, we simplify matters here and interpret basic untensed sentences as properties of times. The reason is that the facts Condoravdi accounts for with this event apparatus are only peripheral for the discussion at hand.

<sup>13</sup>She uses the term *metaphysical* modal base.



$PRES : \lambda P \lambda w. AT(now, w, P),$   
 $PERF : \lambda P \lambda w \lambda t. \exists t' : t' < t \ \& \ AT(t', w, P),$   
 $might/may : \lambda MB \lambda P \lambda w \lambda t. \exists w' \in MB(w)(t) \ \& \ AT([t, -), w', P),$   
 $would/will : \lambda MB \lambda P \lambda w \lambda t. \forall w' \in MB(w)(t) \Rightarrow AT([t, -), w', P),$   
 $AT(t, w, P) : \exists t' : t' \subseteq t \ \& \ P(w)(t')$

(where  $P$  is a property of times,  $MB$  is a time-sensitive accessibility relation, and  $[t, -)$  is the time-interval starting with  $t$  and without right boundary).

$\llbracket (105) \rrbracket = \lambda w. \exists t < now \ \exists w' : [w' \cong_t w \ \& \ AT([t, -), w', Peter \ win \ the \ game)]$

Figure 6.4: Condoravdi's approach to non-root modals

the evaluation time for the property of times in its scope forward, this property might actually refer to the present or the future. However, this approach says nothing about the surface past tense markings on the modalities *might* and *would* and why they appear not to be interpreted as past tenses. Actually, if you look at the trees given in figure 6.4, you see that according to Condoravdi's proposal these modals are not semantically interpreted as bearing a past tense. Instead they are analyzed as carrying present tense.<sup>14</sup> In a handout, Condoravdi (2003) proposes that the past morphology on *might* must (and on *would* can) be interpreted as marking of a subjunctive mood. If the past tense is interpreted this way the tense applied to the modal is the present tense. For the subjunctive mood she proposes that it expresses domain widening, without providing any details on this point. While this is certainly a way to approach the part of the puzzle of the missing interpretation concerning the simple past, it is no longer an approach along the lines of the past-as-past approach. It falls into the second group of theories, past-as-modal approaches, that we will discuss below. Hence Condoravdi (2002, 2003) is a mixed approach. The contribution of the perfect

<sup>14</sup>Condoravdi proposes that *might* is always interpreted as bearing the present tense, while *would* allows also for a past tense reading.

in modal contexts (and in the consequent of conditionals) is explained along the lines of a past-as-past approach, but for the interpretation of the simple past she sketches a past-as-modal proposal.

Until now we have only discussed the puzzle of the missing interpretation in so far as it applies to the consequent of conditionals. The approach of Condoravdi does not say anything about the antecedent. This is not very surprising, given that the proposal is meant to describe the meaning of modals and not of conditionals. Nevertheless, we might try to think of what predictions it could make for the antecedent, if the antecedent is interpreted as a modifier of the modal base. Then, if the simple past occurring in the antecedent of a subjunctive conditional is interpreted as simple past, the meaning of subjunctive conditionals is not correctly described. In this case the modified restrictor of a conditional like (101), here repeated as (106), would consist of those world epistemically or ontically accessible at the utterance time where at some point in the past Peter caught the plane.

(106) If Peter took the plane, he would be in Frankfurt this evening.

This does not capture correctly the intuitions concerning *would* conditionals. We may, however, extend Condoravdi's proposal and suggest that past morphology on non-modal verbs can also be interpreted as selecting a subjunctive mood in the contexts of conditionals. Furthermore, we may propose that in this case the verb is interpreted as marked for the present tense. The predictions made by this approach are already much better, but we cannot account this way for the possibility that the antecedent is interpreted at some time in the future. According to Condoravdi the simple present always refers to the utterance time. The problem could be solved by also extending the evaluation time for the property of times described in the antecedent forward – as Condoravdi (2002) proposes for the evaluation time of the phrase in scope of a modal. But so far, nothing in the approach explains why the extension of the evaluation time in scope of a modal should apply to the antecedent as well. Finally, a word on the interpretation of the syntactic perfect in the antecedent of *would have* conditionals. It is clear that the perfect of antecedents of *would have* conditionals is not involved in the structural ambiguity proposed by Condoravdi. Thus it will be interpreted in situ in the antecedent. This means that the approach so far cannot account for the possibility that antecedents of *would have* conditionals refer to the present or the future.

In the following we will discuss some problems that apply to past-as-past approaches in general and that motivate our choice not to follow this line of approach when it comes to the interpretation of the simple past in conditional sentences. Afterwards, we will also discuss problems for the back-shift interpretation rule all the past as past approaches discussed here adopt. These problems motivate our choice to dismiss a mixed proposal like Condoravdi (2002).

**Problems with accounting for the puzzle of the missing interpretation**

A serious obstacle for past-as-past approaches in general is that similar apparently non-temporal interpretations of the simple past can be observed in quite a number of different constructions of English. Examples are counter-to-fact wishes (107a), complement clauses of a comparison starting with ‘like’ or ‘as if’ (107b), the scope of verbs like ‘suppose’, ‘assume’ (107c), and many other constructions.

- (107) a. I wish I owned a car.  
       b. He behaves like he was sick.  
       c. Suppose she failed the test.  
       d. It’s time we left.

A proponent of the past-as-past approach can in principle react in two ways to these observations. He may defend the past-as-past hypothesis for all occurrences of the simple past with an apparently non-temporal meaning. But this seems a position difficult to maintain given the different structures of the examples. At least for the approaches discussed here it is very difficult to see how this would work. Alternatively, he proposes that, while in subjunctive conditionals the past tense simply means past time reference, something different is going on in all these other cases. The problem, then, is that there is cross-linguistical evidence for a connection between all these apparently non-temporal uses of past tense: there are quite a number of different languages that all show such non-temporal uses of their past tense marker. James (1982) lists 13 languages from different language families that appear to use their past tense marker also non-temporally: English, French, Latin, Classic Greek, Russian, and Old Irish (Indo-European), Cree (Algonquian), Tonga and Haya (Bantu), Chipewyan (Athabaskan), Garo (Tibeto Burman), Nitinaht (Wakashan), and Proto-Uto-Aztecan (in the reconstruction of Steele). Furthermore, these languages employ the marker in similar contexts. All of them use it in certain conditional constructions without it marking past time reference in any obvious way. Many languages share other uses as well. Thus, there seems to be some pattern behind extending the past tense markers to apparently non-temporal uses in conditionals and other constructions that one has to account for.

**Problems with accounting for the meaning of *would have* conditionals**

It is not difficult to see that the back-shift interpretation rule for past conditionals sketched above can be interpreted as an instance of the similarity approach. The similarity relation it gives rise to is not very interesting, though: for instance, for the ontic modal base everything that happens after the closest point where the antecedent was true at some ontic alternatives does not count at all for



similarity, but before that point everything counts. The order relation between worlds that fits this informal description can be defined as follows:  $w_1 \leq_w w_2$  iff  $\forall t : w_2 \cong_t w \Rightarrow w_1 \cong_t w$ . From this perspective the back-shift interpretation rule is just another similarity-based account of the meaning of conditionals that lets the past dominate the similarity relation. We have seen in section 5.3.2 of the previous chapter that there are empirical problems with giving the past priority for similarity in the interpretation of *would have* conditionals. As can be expected, these problems show up here again. One can distinguish two basic assumptions made by the past modality approach to similarity. In the following, both will be shown to lead to false predictions.

**1. Only the past counts.** The back-shift interpretation rule assumes that for similarity only the past of the decision point of the antecedent counts. But example (67), repeated here as (108), shows that facts that are decided after this point can also count for similarity or the truth of *would have* conditionals.

*A coin is going to be thrown and you have bet \$5 on heads. Fortunately, heads comes up and you win. You say*

(108) If I had bet on tails I would have lost.

This *would have* conditional is intuitively true. However, at the moment when you decided to bet on heads, it was still not settled which side of the coin was going to come up (if there is something that should lead to ontic alternatives then it is such a chance event). But that means that at any time before you decided to bet on heads including the time of decision, there are ontic alternatives where head comes up as well as ontic alternatives where tails comes up. Thus, if you go back in time to some point where ontic alternatives are admissible that make the antecedent true, there will always also be ontic alternatives admissible that make the antecedent true and the consequent false. Hence, the theory cannot predict the truth of the *would have* conditional (108).

Ippolito (2003) suggests allowing more than just the past of the decision point of the antecedent to count for the similarity relation. Unfortunately, she does not provide any information on how exactly that should work. Even if she could provide an answer to this question, the next problem that will be discussed cannot be handled this way. Another idea how to deal with the coin example is to let the decision point of the consequent, instead of the decision point of the antecedent, be the moment at which the ontic alternatives have to be checked. While this helps for the problem at hand, this approach cannot deal with the Kennedy example. There the situation is exactly opposite to the coin example. In the coin example the decision point of the antecedent precedes the decision point of the consequent, in the Kennedy example it is exactly the other way around. Additionally, the problem discussed below would also apply to this variation of the

back-shift interpretation rule.

**2. Everything of the past counts.** The back-shift interpretation rule assumes that every bit information about the past of the decision point of the antecedent counts. Example (65), here repeated as (109), shows that this is not true: some aspects of the past may not count.<sup>15</sup>

*A farmer uses the following strategy to turn his sheep into money. First he tries to sell a sheep to his brother. If he doesn't want it, it gets special feeding and some weeks later the farmer tries to sell it to the butcher. If the butcher doesn't want it, he gives it as a gift to the local zoo. One of the sheep is a particular favorite with his little son Tom. Tom doesn't know what became of Bertha, his favorite, because he was away for four weeks. The first thing he does after coming back is run to the zoo. He utters a yell of great relief when he spots his beloved Bertha among the animals there. On request Tom says:*

(109) If Bertha hadn't been here, she would have been at the butcher's.

Intuitively, this sentence is false. However, the back-shift interpretation rule cannot account for this intuition. Let  $t_1$  be the moment when the brother decided not to buy Bertha, and  $t_2$  the moment when the butcher decided not to buy Bertha. For all  $t$  between  $t_1$  and  $t_2$  there are worlds in the ontic modal base where Bertha ends up in the zoo and worlds where Bertha ends up at the butcher, but none where Bertha is bought by the brother. Any of these times  $t$  makes the *would have* conditional (109) true according to the back-shift interpretation rule.

These two examples show that essential basic assumptions underlying the back-shift interpretation rule are wrong: there are events happening after the decision point of the antecedent that may be relevant for the truth conditions of *would have* conditionals, and there are aspects of the past of the decision point that have no impact on the truth conditions. We therefore conclude that the back-shift interpretation rule – independent of how it is derived from the compositional structure of conditionals – is not appropriate to describe the truth conditions of conditionals. This means that all approaches discussed so far have to be dismissed.

### 6.2.3 Past-as-modal approaches

In this section past-as-modal approaches will be discussed. The common element of all these approaches is that they claim that the standard meaning assigned to

---

<sup>15</sup>This example is based on an example from Bennett (2003).

the simple past in English – that is that the simple past<sup>16</sup> refers to a contextually introduced past time – is not correct. Instead, they propose that in some contexts the simple past can rather bear a modal meaning, expressing hypotheticality or distance from reality. Apart from this point of agreement there is a great diversity in how this basic idea is worked out in different past-as-modal approaches. We will distinguish four subtypes of this line of explanation for the puzzle of the missing interpretation: the past-as-unreal hypothesis, the past-as-metaphor hypothesis, the past-as-relict hypothesis, and the life-cycle hypothesis.

### 6.2.3.1 The past-as-unreal hypothesis

The central problem for the past-as-past approach that we have discussed above is that it has difficulties accounting for the generality of the non-temporal uses of the simple past in English: they are not restricted to subjunctive conditionals but can be observed in other constructions as well. One idea for how to account for this observation that immediately suggests itself is that there is a general underlying meaning of all uses of the past tense in English. Thus, according to this position, locating the eventuality described in its scope at some time before the speech time is not the true semantic meaning of the simple past in English. Instead, it has been proposed by different authors that the simple past denotes a much more general and abstract concept that can be described as distance from reality, non-actuality, or hypotheticality. This is what we will call the *Past-as-unreal* hypothesis.

The past-as-unreal hypothesis is the most popular explanation brought forward to account for the missing interpretation miracle. Proponents of the past-as-unreal hypothesis are, for instance, Steele (1975), Langacker (1978), and Palmer (1986). The oldest defender of a past-as-unreal approach may be Joos (1964). One of the approaches best worked out is Iatridou (2000). She proposes that the past tense morpheme in English, or, as she calls it, the *exclusion feature*, *ExclF*, provides a skeleton meaning of the form:

$$T(x) \text{ excludes } C(x).$$

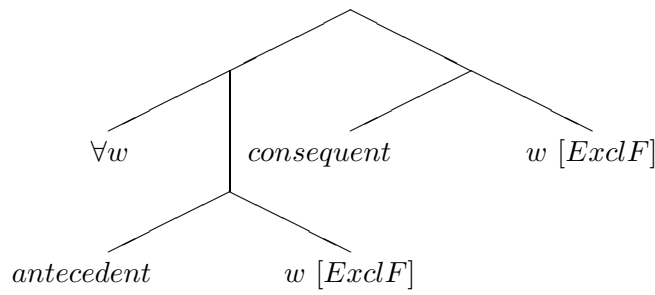
$T(x)$  denotes the object that is currently the topic of discourse, hence, “the  $x$  that we are talking about”.  $C(x)$  stand for “the  $x$  that for all we know is the  $x$  of the speaker”. In this scheme  $x$  can range over different sets of objects, more particularly, it can range over times or worlds. In case the domain consists of temporal intervals, this skeleton comes down to expressing that the topic time lies before the utterance time, which is the time of the speaker.<sup>17</sup> In case the

---

<sup>16</sup>Past-as-modal approaches for the perfect are very rare.

<sup>17</sup>To derive this conclusion, Iatridou has to assume that there are no future times in the domain. It is not entirely clear how she does this. Her argument is that there is no future tense in English, but this does not explain why this means that there is no future time in the domain.

domain is the set of possible worlds,  $\text{ExclF}$  expresses that the topic world is not the actual world, which is the world of the speaker. Iatridou is not very precise in how the meaning of a conditional is exactly calculated. She seems to suggest a logical form where quantification over possible worlds is explicit. In figure 6.5 a possible structure for the semantics of conditionals she might have in mind is sketched. The past tense is interpreted as a feature attached to world-variables that restricts the quantificational domain by excluding the world of the speaker.<sup>18</sup> Iatridou (2001) further proposes that the perfect in *would have* conditionals is interpreted in standard ways. She seems to be unaware of the fact that these conditionals can also refer to the present or the future. Her approach cannot explain this possible interpretation.



(where *antecedent* and *consequent* denote propositions, i.e. functions from worlds to truth values.)

Figure 6.5: Iatridou's approach to conditionals

The past-as-unreal hypothesis is a very attractive approach to the puzzle of the missing interpretation. Firstly, because it is intuitively very appealing. There seems to be something connecting all uses of the simple past, also cross-linguistically. Past time reference locates some eventuality at some non-present time, hypothetical conditionals are often described as those where the speaker at least doubts that the antecedent and consequent actually are true, counterfactual wishes are wishes for something to be true that actually is not true, and for many other uses of the past tense similar paraphrases involving some notion of non-actuality can be given. Secondly, the past-as-unreal hypothesis is attractive because it keeps semantics simple. This is nicely illustrated by the approach of Iatridou (2000), where everything is interpreted in situ, and no manipulations in the surface syntactic structure are involved. Furthermore, and unlike to other past-as-modal approaches we will discuss below, no ambiguity is added to the

<sup>18</sup>Strictly speaking, what is meant by antecedent and consequent in figure 6.5 is the respective phrase without the past tense marking.

lexicon for the simple past. It is just that the meaning is more general than proposed by standard approaches to the simple past.

Nevertheless, this line of approaches has certain drawbacks. First, these approaches provide in general a very unspecific description of the general meaning of the simple past and the way this meaning contributes to the meaning of sentences. In this respect Iatridou (2000) is already an exception. But also this rather specific proposal leaves a lot of questions unanswered. It is, for instance, unclear, (i) how the temporal location is derived in case the exclusive feature applies to worlds, (ii) when the exclusive feature applies to worlds, and (iii) whether it can also apply to other domains. These are shortcomings shared by many past-as-unreal approaches. They do not and often also cannot explain when which specification of the abstract meaning of the past is chosen. This is often said to be made clear by the contexts. But this raises the question, how it is possible that in simple sentences the past is always interpreted temporally, while in subjunctive conditionals, this is never the case.<sup>19</sup> What exactly are the contextual features that are responsible for this clear-cut distinction?

A final problem for past-as-unreal approaches that we want to mention here is that the description they propose for the semantic meaning of the simple past in English is not only often very vague, but also in danger of being too general. As we have seen in the previous section, English is not the only language showing non-temporal uses of its past tense marker. It is rather a phenomenon that can be observed in languages from quite different families. But while there is a certain similarity between the contexts in which these languages employ this marker, there are also language specific differences. In order to account for the general meaning of the simple past in English a proponent of the past-as-unreal hypothesis has to give a description of this semantic property that singles out those and only those uses made of the Simple Past in English. This is clearly something notions like ‘distance from reality’ and ‘non-actuality’ etc. cannot achieve. The question is, can we do better than this: can we give (for every language that shows the missing interpretation miracle for their marker of past time reference) a general description that selects exactly those uses made of the past tense marker in this specific language. Some linguists (for instance, James (1982)) have serious doubts on this point. This problem is made worse by the fact that proponents of the Past-as-unreal hypothesis do not seem to be aware of it. They often go as far as defending, more or less clearly, the opinion that distance from reality is a kind of universal concept that in some languages is encoded in what is normally analyzed as a past tense marker. How can this account for the subtle differences in the

---

<sup>19</sup>Iatridou (2000) claims, that in conditionals both interpretations of the exclusive feature are possible, but this is not correct. A sentence like (110) can never be interpreted as referring to the past.

(110) If Peter got the plane, we would make it in time.

uses of the past time marker in different languages?

### 6.2.3.2 The past-as-metaphor hypothesis

As we have seen in the last section, the past-as-unreal hypothesis has to face some serious drawbacks. That makes it attractive to look for alternative explanations. However, its central idea, that there is some underlying concept that connects all these different uses of a past tense marker, is intuitively very attractive. The question is: can we keep this aspect of the past-as-unreal hypothesis but nevertheless account for observations like the language specific differences in non-temporal uses of a past tense marker? A solution to this problem has been brought forward in what we will call the *past-as-metaphor* hypothesis, defended, for instance, by James (1982) and Fleischman (1989). According to this position, the non-temporal uses of a past tense marker are extensions of the basic temporal meaning that emerges because of conceptual similarities between locating eventualities in the past and other concepts that speakers want to describe. Hence, a past tense marker can be used as a kind of metaphor. This time the marker of past time reference does not mean the same in all of its occurrences, but there are some conceptual consonances that made it an appropriate metaphor for all of its non-temporal uses. It should not come as a surprise that these conceptual consonances are again described as distance from reality or non-actuality.

This hypothesis, as much as the past-as-unreal hypothesis, can account for the fact that quite a number of different languages show the same pattern of extending the use of their past tense marker – they all employ the same metaphor. But it can also explain the language-specific differences: of course, different languages may conventionalize different uses of this metaphor. Furthermore, the dominance of the temporal reading of a past tense marker as well as the observation that, in contrast to non-temporal uses, this reading does not have to be specified by the context, follows immediately from the fact that the temporal meaning is the basic meaning of a past tense marker.

But also the Past-as-metaphor hypothesis raises some as yet unresolved questions. For one thing, as was the case for approaches adopting the past-as-unreal hypothesis, theories following the past-as-metaphor hypothesis are not very specific on what meaning a past tense marker has in its non-temporal uses. Furthermore, one might expect that if a past tense marker is so often used as distance from reality metaphor, then other parts of the language, such as, for instance, future tense markers or spatial expressions are at least in some languages employed as distance metaphors as well. Fleischman (1989) claims that this is true for the future, but she has to admit that this holds to a much less degree. To the knowledge of the author there is no language that employs spacial expressions as (grammaticalized) metaphor for hypotheticality. But more serious cross-linguistical studies may prove the author wrong on this point. Another potential problem of the past-as-metaphor hypothesis is that it has difficulties accounting

for certain regularities in the extension of past markers to non-temporal uses. In her study of 13 languages that show such non temporal uses of a past tense marker James (1982) observes that all the languages use their past tense marker in the consequent of counterfactuals referring to the past, the present or the future. A majority of the languages also use the past in the antecedent of such conditionals and in counterfactual wishes. Still many languages have it in conditionals that are not counterfactual, but where intuitively the speaker takes it to be unlikely that antecedent and consequent turn out to be true. Other uses of the past tense marker become more and more language specific. Hence, the metaphor seems to be more likely to be applied to some contexts than to others and the question is why this should be the case. James (1982) proposes that the likeliness depends on how strong the distance from reality is that is expressed by a certain context. According to her the consequent of counterfactual is the situation most distant from reality, while already the antecedent of such a conditional is a bit less distant, and so forth. Such an ordering needs serious motivation – at least more serious than what is provided by James (1982). Furthermore, we still miss an explanation for why distance from reality decides how likely the past tense metaphor is conventionalized in a context.

### 6.2.3.3 The past-as-relict hypothesis

A line of explanation of the puzzle of the missing interpretation that is more inspired by the diachronic facts about English than by the cross-typological observations is the *past-as-relict* hypothesis. According to this hypothesis, the past indeed means something else in those constructions where its temporal meaning appears to be missing. In these cases it carries the meaning of the old English past subjunctive complex that became indistinguishable from the past indicative in Middle English. At this stage the past indicative starts carrying two meanings, the standard temporal meaning, traditionally conveyed by the past indicative, and the hypothetical meaning of the past subjunctive. Such a position has been suggested, for instance, by James (1986) and Dahl (1997).

This approach nicely fits the diachronic data concerning English. In contrast to the last two proposals we have discussed, the past-as-relict hypothesis does not depend on any conceptual similarity between the temporal meaning of a past tense marker and the meaning it carries, for instance, in subjunctive conditionals. Instead, it is claimed that independently motivated diachronical changes forced the simple past to adopt a second meaning and to become ambiguous. This makes it more difficult for the past-as-relict hypothesis to explain why so many languages show similar non-temporal uses of their past tense markers. Do they all share the same history with respect to their past and subjunctive marker as English? If yes, what is the motivation for this cross-linguistic diachronic process?

There are also other intriguing questions for the past-as-relict hypothesis. For instance, it has been observed that already in Old English, before the subjunc-

tive/indicative distinction disappeared for the past, conditionals with the past subjunctive referred to the past as well as to the present or the future. Hence, already at this stage the past appears to have lost its temporal meaning. How is this to be explained? To start with, what was the meaning of the past subjunctive complex in Old English and is the meaning of Past, hence, in conditionals today? James (1986) describes this again very vaguely as distance from reality. This certainly needs to be made more precise.

Let us discuss one final problem for the past-as-relict hypothesis. To account for the non-temporal uses of the simple past in other contexts besides subjunctive conditionals, proponents of this hypothesis would probably propose that also in these contexts the past subjunctive was used in Old English and then in Middle English this function was taken over by the simple past. According to James (1986) in these contexts and many others Old English indeed uses the subjunctive but not necessarily the past subjunctive. Instead, James (186) provides a number of examples where past and present still show a normal temporal meaning when combined with the subjunctive mood. If James (1986) is correct<sup>20</sup>, then this casts some doubt on the idea that the simple past simply adopted the meaning of the past subjunctive. There is another, similar observation supporting the conclusion that there is a difference between the uses of the past subjunctive in Old English and the uses of the past in Contemporary English. According to Visser (1973), in Old English only counterfactuals were marked by the past subjunctive, while nowadays the past does not necessarily convey counterfactuality in conditionals.<sup>21</sup>

#### 6.2.3.4 The life-cycle hypothesis

In the same paper cited above Dahl (1997) also makes a second proposal for the meaning of past in English, that is independent of the past-as-relict hypothesis. In section 6 of his paper he proposes a diachronic life-cycle for a marker of counterfactual constructions, that is intended to apply cross-linguistically. According to this life cycle, past tense markers systematically develop through four stages a use as marker of hypothetical constructions. We quote Dahl's description of the four stages (Dahl 1997: 109).<sup>22</sup>

- (1) "In the first stage, the marker would be (a) restrained to past reference, (b) imply counterfactuality in the strict sense (dependence on a condition known to be false), (c) be optional.

---

<sup>20</sup>His position stands in conflict with Visser (1963), who claims that in those contexts where now the past occurs in its non-temporal meaning, in Old English the past subjunctive was used.

<sup>21</sup>What is meant by counterfactuals differs highly between authors and also with respect to one and the same author. In this case the risk that there is a different understanding of the word 'counterfactual' is rather small because Visser explicitly describes these conditionals as 'hypothetical period with unrealizable or unreal antecedent' (Visser, 1963: paragraph 861).

<sup>22</sup>Dahl emphasizes that this proposal is only incompletely supported by the data and based on observations concerning quite different languages.



- (2) In the second stage, the marker would become obligatory in past counterfactual contexts.
- (3) Then, the constraints on its use would be gradually relaxed. The first thing to go would be the temporal condition [...].
- (4) Once the construction has become possible with non-past reference, the risk that the counterfactuality constraint is also relaxed will be imminent.”

The described diachronic development Dahl (1997) sees particularly clearly exemplified in how the perfect auxiliary *is* and *was* was used in conditional constructions in different Germanic languages. It is well known that the perfect developed only at a later stage of English and German and so did counterfactual pluperfects. It is also supported by data that the pluperfect was added to a system where time reference was not marked in subjunctive conditionals. At this stage the perfect conveyed past reference and was, according to Dahl (1997), not obligatory (as it still is in Bulgarian).<sup>23</sup> Later on, Dahl proposes, the past perfect became an obligatory marker of past counterfactual conditionals – thus, moves to stage two of the life cycle. Dahl (1997) claims, referring to a similar statement of Jespersen (1924), that the use of the past perfect in conditionals referring to the present or the future is only a recent development. He takes this as evidence that the perfect in English just moved from stage 2 to stage 3.

Dahl (1997) seems to defend the position that the meaning of the simple past in English also developed along this cycle. This appears to be in conflict with the past-as-relict hypothesis that he defends in earlier sections of the same paper. According to this theory, the past obtained its hypothetical meaning because the form originally encoding hypotheticality, the past subjunctive, got lost. However, Dahl (1997) also admits that Germanic subjunctive conditionals<sup>24</sup> are complex constructions consisting of different elements that interact and whose histories may also influence each other. Such interactions may then be responsible for why the past did not develop straight along the cycle Dahl proposes, but took over the meaning of the past subjunctive complex. A different story one could think of is that as much as past and past perfect are different markers of hypotheticality this is also the case for past and the subjunctive in Old English. For instance, one could propose that the simple past was at this time already in stage 3 of the life cycle and conveyed, independent of temporal reference, counterfactuality in conditional sentences. The subjunctive, on the other hand, was a very general hypotheticality marker. After the subjunctive for past forms disappeared the

---

<sup>23</sup>Dahl suggests that the driving factor to introduce the perfect was the need to have a way to emphasize the counterfactuality of the proposition.

<sup>24</sup>Dahl (1997) actually uses the word ‘counterfactuals’. Given the general use he makes of this term, I think that he means with it roughly the same as we do with ‘hypothetical conditionals’.

past developed from stage 3 to stage 4 and took over some of the functions of the subjunctive.<sup>25</sup>

A clear point in favor of Dahl's (1997) life-circle hypothesis is that it can explain to a certain extent the diachronic data on the changes in form and meaning of English conditionals sentences. A second advantage is that it suggests an explanation for James' (1982) observation that in all languages with a hypothetical past tense markers it is used in past counterfactuals. According to Dahl, past counterfactuals are the context where past tense first develops a non-temporal meaning. Other uses follow when the marker of hypotheticality reaches stage 4 of his life circle. But the approach also makes a lot of strong predictions that should first be verified. For instance, it has to be checked whether the past perfect in English really did follow the different stages of the life circle Dahl has proposed. Furthermore, it still has to be verified, in how far this is a general cross-linguistically correct description of how past tense markers develop a hypothetical meaning. Finally, the proposal also leaves important questions unanswered. For instance, Dahl's description of the meaning of a hypotheticality marker in stage four is very vague. In consequence, we still do not know what the simple past in subjunctive conditionals means. Second, we miss an explanation for why a past tense marker starts to imply counterfactuality in stage 1 of the circle.

To summarize the discussion of this section, the literature of the past-as-modal approaches is characterized by a lot of interesting ideas. However, only rarely are these ideas developed into concrete proposals. This makes it difficult to evaluate them. In section 6.4 we will develop a new approach to the meaning of the simple past and the perfect in English and propose an explanation for the puzzle of the missing interpretation. This approach will follow the idea of the past-as-modal approaches and propose a non-standard meaning for the simple past in contemporary English. We hope that the precise formulation of this new approach will lead to a more elaborate discussion on the synchronic, diachronic, and typological questions the puzzle of the missing interpretation raises.

## 6.3 The puzzle of the shifted temporal perspective

In this section we will have a closer look at the puzzle of the shifted temporal perspective. In contrast to the puzzle of the missing interpretation, it concerns in the first instance indicative conditionals. Most students of the semantics of English conditionals take the case of indicative conditionals to be simple compared with subjunctive conditionals. The general opinion is that in this case we can use

---

<sup>25</sup>Dahl (1997) does not seem to claim that all markers of hypotheticality, and hence also the old Subjunctive in English, developed out of a past tense marker.

something like an analysis as strict conditional and interpreted all tenses *in situ*. But also for indicative conditionals we observe that such an approach, combined with standard proposals for the meaning of English tenses, does not always make the right predictions.

We are interested in one particular observation that disturbs this simple picture. It turns out that in the consequent of (indicative) conditionals – as well as in the scope of modals – the reference time<sup>26</sup> for the interpretation of the tenses, which is normally assumed to be the utterance time, can shift to the future. This is what we will call the *puzzle of the shifted temporal perspective*. In this section we will study the relevant data that constitute the puzzle and discuss some approaches trying to account for them. But first we will discuss a related, well-known observation: the evaluation time for present tense sentences and the phrase in the scope of modals can also be shifted to the future. A clear view on this phenomenon is needed to develop a proper understanding of the puzzle of the shifted temporal perspective discussed afterwards. In the discussion of the data we will rely heavily on observations made by Crouch (1993), which represents by far the most extensive work on the puzzle of the shifted temporal perspective to date.

### 6.3.1 The observations

**Future-shifted evaluation times.** The first observation that will be discussed here is commonly known and has already been mentioned in the preceding section. The observation is that the evaluation time of present tense sentences – and something similar holds for the phrase in scope of modals – can lie in the future. Let us start with the case of the simple present. It has often been observed that, even though normally the evaluation time of sentences in the simple present, in many cases is the utterance time and an evaluation in the future is not possible (see (111a) and (111b)), there are exceptions to this rule (see (111c) and (111d)).

- (111) a. \*I come to your party tomorrow.  
       b. I will come to your party tomorrow.  
       c. The train arrives tomorrow at 9 pm.  
       d. Arsenal plays Spurs at home next week. (Crouch, 1993: 33)

This has lead various linguists to propose that the simple present localizes the evaluation time in the non-past instead of at the utterance time (Lyons 1977,

---

<sup>26</sup>Recall from Chapter 4 that in this book *reference time* refers to the time with respect to which tenses localize the evaluation time of the phrase in their scope. This time normally equals the utterance time. But as we will see in this section, sometimes it can also be a time in the future of the utterance time.

Nerbonne 1985, Comrie 1985, Kaufmann 2005). This would immediately account for future uses of the simple present in simple sentences. The challenge for such an approach is to explain why many future uses of the simple present are semantically anomalous (see (111a)) and a reformulation with *will* has to be used (see (111b)). The explanation seems to lie in the observation that future uses of the present tense in simple sentences are only acceptable if the fact in the scope of the present tense is interpreted as already settled at the utterance time, or, as Kaufmann (2005) puts it, sentences using the simple present to refer to the future come with a *certainty condition* (see (111c) and (111d), similar observations have been made by Lakoff 1971, Goodman 1973, Quirk et al. 1985, Comrie 1985 and many others).

The picture is further complicated by the observation that in indicative conditionals future evaluation times for present tense antecedents is more the rule than the exception. Furthermore, these future uses do not come with a certainty condition. An antecedent as in (112) is not interpreted as selecting those worlds in which it is certain at the utterance time that the bimetallic strip will bent. In general, Crouch (1993) observes that there appear to be no restrictions on when a present tense antecedent can take on a futurate interpretation. Accounting for this observation is one of the challenges compositional approaches to the meaning of indicative conditionals have to face.<sup>27</sup>

(112) If the bimetallic strip bends, the temperature rises.

A similar future-shift of the evaluation time can be observed for phrases in the scope of modals: the evaluation time of such a phrase can be localized in the present or in the future. Reference to past eventualities is only possible if the perfect is used. For instance, example (113a) can just as well describe the location of John at the utterance time as at some time in the future of the utterance time. However, it cannot refer to the location of John in the past. Example (113b) can describe John's whereabouts at some time in the past as well as in the future.

- (113) a. John may/must/will be in London. (Crouch 1993: 42)
- b. John may/must/will have finished the essay (by next Tuesday). (Crouch. 1993: 42)

An important difference between modal contexts and present tense sentences is that for modals the future-shift of the evaluation time is relative to the evaluation time of the modal and not to the utterance time. This fact can easily be ignored, because the evaluation time of a modal very often equals the utterance time. But there are exceptions. First, at least some modals in English still have a past tense

---

<sup>27</sup>The approach introduced in section 6.4 can deal with this observation. But this is not the puzzle of the shifted temporal perspective.

form that is interpreted as locating the evaluation time of the modal in the past. An example is *would*. In example (114) the evaluation time of the property in the scope of the modal, x's being a king, is not localized at or in the future of the utterance time, but just has to follow the time when the child was born, which is the evaluation time for the modal.

(114) A child was born that would be king.

Second, there are contexts in which a present tense modal is evaluated in the future, as in (115). In this case the evaluation time of the phrase in the scope of the modal can lie at this future time or in its future, but not at some time following the utterance time and preceding the evaluation time of the modal. My permission to drive a car becomes effective tomorrow when I pass the examination. This permission does not legalize any future driving taking place before the time when I pass the examination.<sup>28</sup>

(115) If I pass the examination tomorrow, I can drive a car.

In an important respect the future-shift of the evaluation time in the scope of modals behaves similar to future uses of the present tense in the antecedent of indicative conditionals. We observe that reference to the future in modal contexts is not bound to the certainty condition. That means that the (normal) meaning of a sentence like (113a) is not that in some/all possible world(s) in the relevant modal bases it is already determined at the utterance time that John is in London at some future time. In this respect the meaning of a simple sentence with present tense like (111a) differs from a sentence using *will* like (111b).

**Future-shifted reference times.** After this excursion to future-shifted evaluation times, we now come to the observations that constitute the puzzle of the shifted temporal perspective. The relevant observation is that in the consequent of conditionals and in modal contexts the reference time of tenses can be shifted to the future. We will discuss this puzzle step-wise for different constructions. We start with consequents of indicative conditionals that do not contain a modal (section A). Then we turn to consequents of indicative conditionals that do contain a modal (section B), then to consequents of subjunctive conditionals (section C). Finally, we will consider relative clauses in the scope of modals (section D).

**(A)** We start with indicative conditionals without a modal in the consequent. If both antecedent and consequent of such a conditional stand in the present tense, then the consequent is evaluated at a time overlapping or following the evaluation time of the antecedent. Sentence (116) cannot be interpreted as saying that

---

<sup>28</sup>The intuitions for the German and Dutch counterparts of example (114) differ. In these languages one would use a different modal to express granting a permission in this context.

the bending of the bimetallic strip tells us that the temperature must have risen beforehand. It can only mean that the temperature rises in consequence of the bending of the strip.

- (116) If the bimetallic strip bends, then the temperature rises. (Crouch, 1993: 1)

This is unexpected. One would think that *the temperature rises* either refers to the utterance time, or, if one adopts an interpretation of the simple present that allows future reference, any time after the utterance time. In particular, one would expect that times between the utterance time and the evaluation time of the antecedent provide possible evaluation times for the consequent. But (normally)<sup>29</sup> sentences like (116) come with the inference that these times are excluded. A straightforward explanation of this observation is that the reference time of the present tense in the consequent is the evaluation time of the antecedent and not to the utterance time. That would immediately account for the observation. But why should that be the case? Furthermore, if this analysis is correct, then the interpretation of the present tense cannot be deictic.

We make exactly the same observation for past tense consequents without a modal. If a present tense antecedent that refers to the future is followed by a past consequent (without modal), then the consequent can be evaluated at some time after the utterance time of the conditional and before the evaluation time of the antecedent. This is illustrated with example (117a). The interview may very well take place after the sentence is uttered. Again, this is surprising. Given standard interpretations of the simple past, one would expect that the consequent is interpreted at some time before the utterance time.

- (117) a. If he comes out smiling, the interview went well.

The reference point for tenses in the consequent of a conditional is not shifted, if the antecedent stands in the simple past. If a past tense antecedent is followed by a present tense consequent, the consequence is not evaluated at the evaluation time of the antecedent or any other past time following the evaluation time of the antecedent. Instead, the evaluation time is the utterance time, as expected (see example (118a)).

- (118) a. If John had a packet of cigarettes in his pocket, then he smokes. (Crouch, 1993: 36)
- b. If the bimetallic strip bent, then the temperature rose. (Crouch, 1993: 1)

---

<sup>29</sup>We will come to some exceptions below.

Furthermore, if the consequent uses the simple past as well, then it does not have to be evaluated at some point before the past evaluation time of the antecedent, but can be interpreted at any point before the utterance time of the conditional. Thus, we observe no shift for the reference time of the tense in the consequent. This is illustrated with example (118b). This example has two readings. Either one interprets the sentence as saying that the temperature rose as consequence of the bending of the bimetallic strip – in which case the past evaluation time of the antecedent would precede the evaluation time of the consequent. Or the sentence is understood as saying that the bending of the bimetallic strip tells us that the temperature rose. In this case the evaluation time of the consequent precedes the evaluation time of the antecedent. In Chapter 5 we called conditionals with this order of the evaluation time of antecedent and consequent *backtracking conditionals*. The first reading would be excluded, if the reference time of the second past tense was shifted to the evaluation time of the antecedent.

It is important to realize that the future-shift of the reference time of tenses in conditionals is not obligatory. One can find examples that follow the interpretation predicted by the standard semantics for the tenses.<sup>30</sup> This is illustrated by the next two examples. Both conditionals contain the same antecedent evaluated at some future time. Furthermore, both conditionals make in the consequent the same claim about some future time in the past of the evaluation time of the antecedent. The first sentence, however, uses a past tense in the consequent to express this conclusion. This means that in this case the reference time of the tense cannot be the utterance time, but has to lie in the future of this time. More precisely, its reference time appears to be set to the evaluation time of the antecedent. The second sentence uses a present tense for the same conclusion. Now, the reference time of the tense cannot be shifted to the future evaluation time of the antecedent, otherwise we have to allow for past uses of the present tense. But, as Crouch observes, the reading without a future-shift of the reference time is only possible if the antecedent comes with the certainty condition. This observation will be crucial for our account for this reading.

- (119) a. If the train arrives tomorrow at 9 pm, then it left Sidney yesterday morning 10 am.
- b. If the train arrives tomorrow at 9 pm, then it leaves Sidney at 3 pm this afternoon.

---

<sup>30</sup>Crouch (1993) seems to argue that with respect to this point the simple past and the simple present behave differently: while the present tense allows for the reading referring to the utterance time of the conditional or modal the same is not true for the past tense. The data do not support his point of view. While they are in principle consistent with his claim, they are as well consistent with the position that both tenses show the relevant ambiguity. Furthermore, intuitively in an example like (119a) the past is evaluated with respect to the utterance time.

(B) So far we have only discussed the puzzle of the shifted temporal perspective for indicative conditionals without a modal in the consequent. Do the observations mentioned above also hold in case there is a modal present? Before we answer this question, first notice that in the consequent of conditionals the relation between the evaluation time of the modal and the evaluation time of the phrase in its scope is the same as in simple modalized sentences. That means that the evaluation time of the property in scope of the modal lies at the evaluation time of the modal or in the future of this time, except for combinations of the modal with a perfect. Keeping this in mind, we see that the puzzle of the shifted temporal perspective shows up in modalized indicative conditionals as well. That means that if a present tense antecedent refers to the future, then the present tense modal in the consequent (normally) cannot be evaluated at some present or future time before the evaluation time of the antecedent. Hence, it seems as if tenses in the consequence refer to the evaluation time of the antecedent instead of the utterance time. This is, again, not the case if the antecedent stands in the past tense. For illustration see the examples (120a) and (120b) below. The first example shows that the evaluation time of the modal can be shifted forward to a future evaluation time of the antecedent. The second example illustrates that the same is not the case for past antecedents.

- (120) a. If the strip bends, the temperature may rise. (Crouch, 1993: 45)
- b. If the bimetallic strip bent, the temperature will/may/must rise. (Crouch, 1993: 45)

Also for modalized consequents of indicative conditionals the future-shift of the reference time is not obligatory. It is possible that the evaluation time of a present tense modal in the consequent is not shifted to the future evaluation time of a present tense antecedent. However, as for unmodalized consequents this appears to be restricted to cases where the antecedent is interpreted as settled or, in Kaufmann's (2005) words, comes with the certainty condition. We again illustrate this point with some examples. The first sentence (121a) is hardly acceptable. It is only interpretable in a context where the antecedent can be read as predetermined (which is very difficult to do for the context of job interviews). To express that from John's smiling we can conclude that the interview went well, the use of the past tense or a present perfect is necessary (see (121b) and (121c)). The fourth sentence (121d) illustrates that the modal can be evaluated at the utterance time in case the antecedent can be read as already settled.

- (121) a. \*If John comes out smiling, the interview will go well.
- b. If John comes out smiling, the interview went well.
- c. If he comes out smiling, the interview has gone well.



- d. If this train arrives only at 9 pm tomorrow, Mary will buy tickets for another one.

So far modalized consequents seem to behave exactly like non-modalized consequents of indicative conditionals. But there is also a difference in their temporal properties. A present tense modal in the consequent can never be evaluated in the future of the evaluation time of a present tense antecedent. Thus, it is either evaluated at the utterance time or at the evaluation time of the antecedent. This restriction does not apply to indicative conditionals without a modal in the consequent. In sentence (116) the evaluation time of the consequence lies in the future of the evaluation time of the antecedent. Compare this with sentence (120a). The evaluation time of the modal *may* cannot lie in the future of the evaluation time of the antecedent *if the strip bends*. As Crouch (1993) puts it, the possibility of a rise in temperature comes into being as soon as the antecedent event occurs.

(C) After this extensive discussion of indicative conditionals, one may wonder whether the puzzle of the shifted temporal perspective occurs with subjunctive conditionals as well. Indeed, for *would* conditionals we make exactly the same observations as for indicative conditionals with present tense antecedent and a present tense modal in the consequent.<sup>31</sup> If the antecedent is evaluated at some future time, then the modal in the consequent is normally evaluated at this future time as well (see (122a)). It can also refer to the utterance time of the conditional, but in this case the antecedent comes with the certainty condition (see example (122b)). If the speaker wants to locate the consequent at some time before the evaluation time of the antecedent, but the antecedent cannot be assumed to be settled, then the perfect has to be used (see (122c)). If a past perfect antecedent is combined with a *would* consequent, then the antecedent refers to the past and the consequent to the utterance time (see (122d)).

- (122) a. ?If he came out smiling, the interview would go well.  
 b. If this train arrived only at 9 pm tomorrow, Mary would buy tickets for another one.  
 c. If he came out smiling, the interview would have gone well.  
 d. If she had married Cliff, she would live in the States now.

The temporal properties of *would have* conditionals are very difficult to access. Observations on the temporal relation between antecedent and consequent are sensitive to the reading applied to the conditional and blurred by the many possible temporal interpretations for antecedent and consequent. It seems to be the case that if both the antecedent and the phrase in scope of the modal in the

---

<sup>31</sup>This is not discussed in Crouch (1993).

consequent are evaluated in the past, the evaluation time of the consequent can lie before the evaluation time of the antecedent, be simultaneous or lie in the future of the evaluation time of the antecedent (see also the discussion on backtracking in section 5.3.1 in the previous chapter). For *would have* conditionals with antecedent referring to the present or the future, more empirical investigations are needed before one can make any definite claims, particularly, on the question whether the puzzle of the shifted temporal perspective applies here as well.

(D) Let us finally point out that modal contexts on their own also appear to shift the temporal perspective for the phrase in the scope of the modal. While it has been argued that there is no tense immediately under a modal (see Condoravdi 2002 for discussion), tenses can occur in subordinated relative sentences. In this case one makes similar observations with respect to a future-shift of the reference time as we have made for the consequent of present tense indicative conditionals. See, for instance, the following examples.

- (123) a. By 1998, everybody will know someone who died of AIDS. (Crouch, 1993: 2)
- b. Next week, you must show me a problem that you solved on your own. (Crouch, 1993: 44)

The statement (123a), uttered in 1993, expresses that the property described in scope of the modal holds at some point in the future, more particularly 1998. The relative clause *who died of AIDS* refers to the past of this future reference point, but not necessarily to the past of the utterance time. Also for present tense modals Crouch (1993) argues that the present tense in relative clauses can just as well be interpreted as referring to the utterance time of the modal statement. To illustrate this point he uses the examples (124a) and (124b) given below. The problem with these examples is that nothing forces the evaluation time of the modal to be shifted to the future. But this is essential for the examples to underpin the claim Crouch makes. We leave it open for future research, whether indeed he is right on this point, and also whether again the certainty condition accompanies the reading where the evaluation time of the phrase in scope of the modal is localized with respect to the utterance time.

- (124) a. One day I will marry someone who gets rich quick. (Crouch, 1993: 44)
- b. One day I will marry someone who got rich quick. (Crouch, 1993: 44)

### 6.3.2 Approaches to the observations

In the following we will discuss some of the proposals made in the literature to account for the observations reviewed in this section. We will start with the

observed future-shifts of the evaluation time and afterwards come to the puzzle of the shifted temporal perspective that concerns future-shifts of the reference time. The discussion will be relatively short compared to the grade of elaboration of the theories that we will consider. For more details the interested reader is referred to the original literature.

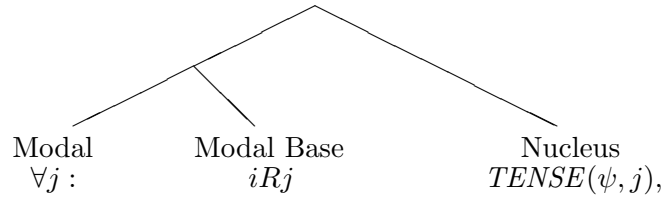
The first observation, discussed in the section *Future-shifted evaluation times*, concerning the future-shift of the evaluation time of present tense sentences and the phrase in scope of a modal, is quite generally acknowledged in the literature. Unfortunately, that does not mean that there exist equally generally accepted accounts for it. It is still an ongoing debate whether the simple present should be analyzed as localizing the evaluation time of the phrase in its scope at the utterance time, or rather at some non-past time. The first position has problems to account for the observation that there exist sentences in the simple present that do refer to the future. To solve this problem without having to give up the position that the present tense always refers to the utterance time, some linguists have proposed that there is sometimes a hidden future operator working in scope of the present (see, for instance, Dowty 1979, von Stechow 2005). But then one has to explain why this operator occurs only in certain circumstances, i.e. when the described fact about the future is assumed to be already settled. To follow the second option and analyze the simple present as locating the evaluation time in the non-past appears to be the more elegant solution. But also in this case it has to be explained why simple present sentences about the future are only acceptable when interpreted as determined at the utterance time.

Kaufmann (2005) takes the second option sketched above. To explain the presence of the certainty condition, he proposes that the semantic present tense stands in scope of an epistemic/ontic modal operator.<sup>32</sup> This operator is hidden in sentences with a bare present tense, but can be realized as an explicit modal like *will*. If the operator is not expressed, it is always the all-quantor. To treat all tenses on a par, Kaufmann (2005) proposes that the simple past also obligatorily stands in the scope of a modal operator. The semantics for modals Kaufmann (2005) assumes comes down to the approach of Kratzer (1979, 1981): they are interpreted as quantifiers over set of indices. An index is a tuple consisting of a possible world and a time. The first argument of the quantor is the restrictor or the modal base, the second argument is the nucleus, described by the property in scope of the modal. The semantic structure Kaufmann suggests for simple sentences with a bare tense appears as presented in figure 6.6.<sup>33</sup>

---

<sup>32</sup>The meaning of these operators works along the lines sketched in section 6.2.2.

<sup>33</sup>This representation simplifies Kaufmann's (2005) approach considerably. For instance, Kaufmann also uses a speech time index that is projected through the whole modal construction to account for the interpretation of deictic elements in modal contexts as temporal adverbials, but not for the interpretation of the tenses. Furthermore, in his approach the modal combines first with the tense and then with the modal base. But we are here not so much interested in how exactly the parts of the construction combine, but rather in the resulting meaning predicted



(where  $i$  and  $j$  are indices,  $R$  is an accessibility relation between indices, and  $\psi$  is a property of times.)

Figure 6.6: Kaufmann's approach to simple sentences

In this representation  $\psi$  is the property in the scope of the tense (or the tensed modal, if overtly expressed). If the tense is the present tense then  $TENSE(\psi, j)$  is interpreted as the claim that  $\psi$  is true at some time at or after the time-component of the index  $j$ . The meaning of the past tense is defined analogously. Figure 6.6 illustrates three important claims of Kaufmann's analysis. First, the tense is interpreted in the scope of the modal operator, in particular, in its nucleus. This is also the case for sentences with an overt modal, where the relevant tense syntactically applies to the modal. Second, this approach also assumes that the modal needs an evaluation index. In figure 6.6 this index is represented by the letter  $i$ . However, the tense present on the modal does not restrict the location of the time-component of this index, as one might expect. As Kaufmann (2005) explains, a matrix sentence operator assures that in simple sentences this index is equated with the speech index. This, of course, only applies if the modal is the highest verb in a matrix sentence. As we will see, in the antecedent of conditionals the evaluation index of the modal can differ from the utterance time. Third, and most surprising, Kaufmann (2005) proposes that tense is not evaluated with respect to the utterance time. It is in its core not at all indexical. Tense is interpreted with respect to the index of the modal base of the modal operator in whose scope it is interpreted. The reference time can be the utterance time, but only if (i) all indices in the modal base share their time-component with the evaluation index of the modal, and (ii) this later index is set to the utterance time. We have already mentioned that the second condition (ii) is sometimes violated. As we will see, in the context of conditionals also (i) does not hold.

Before we come to Kaufmann's (2005) treatment of conditional sentences, let us first point out that this approach can account for the observation that future uses of the simple present assume settledness. This is predicted if the hidden modal operator quantifies over ontic alternatives. In this case a sentence like

---

for the whole sentence.

(111a) is interpreted as claiming that in all possible futures of the actual world I will be at your party tomorrow.

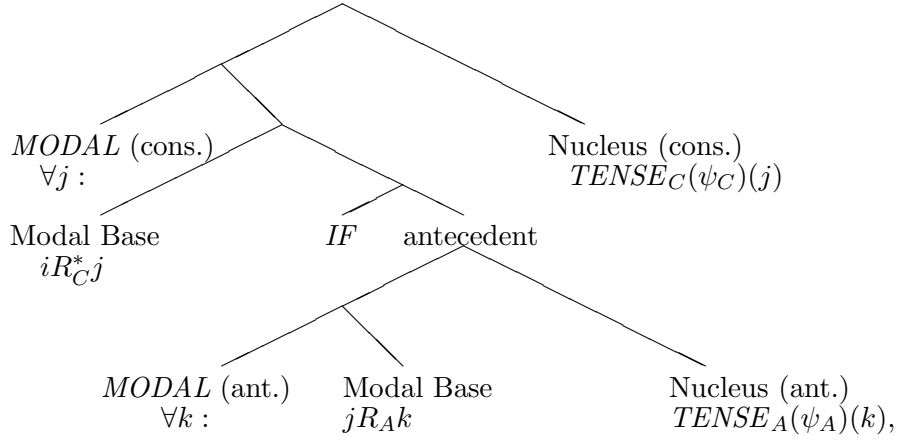
So far we have only discussed some basic claims of the approach of Kaufmann (2005). It goes much further and can actually account for the observation that the certainty condition is absent in most present tense antecedents of conditionals. Let us take a look at the semantics Kaufmann proposes for conditionals. The semantic structure assumed for these constructions looks roughly as illustrated in figure 6.7. Kaufmann assumes that in conditionals the part *if antecedent* modifies the modal base of the modal in the consequent. The new modal base that results from this modification is the set of indexes that are (i) accessible from the evaluation index of the modal or temporally follow indexes accessible from the evaluation index, and (ii) make the antecedent true. The first condition extends the time component of the selected indices forward.

Kaufmann's (2005) theory can explain the observation that the simple present in the antecedent of conditionals does not (necessarily) come with the certainty condition. The evaluation index for the modal in the antecedent ( $j$ ) is set by the modal base of the modal in the consequent  $\{j \mid iR_C^*j\}$ . Because the modal base of the consequent is extended to the future, the time component  $t_j$  of the indices  $j$  in this base can lie in the future of the speech time. The modal base of the antecedent  $\{k \mid jR_A^*k\}$ , and thereby also its tense, is evaluated with respect to this time. Therefore, even though the antecedent claims that the property  $\psi_A$  is determined at time  $t_k$ , this time  $t_k$  may lie in the future of the utterance time. That  $\psi_A$  is settled is thus only claimed for some point in the future, not the utterance time. One may wonder, however, whether it is adequate to describe the meaning of conditionals as asserting the truth of the consequent conditional on the future predetermination of the antecedent.<sup>34</sup>

We have also observed that a similar extension of the evaluation to the future occurs in the scope of modals. However, this time the reference point is not the utterance time but the evaluation time of the modal. To explain this observation, some linguists have suggested that in the scope of modals a hidden future operator or a hidden present tense with possible future reference applies. However, there are convincing arguments speaking against the assumption of tense operators in the scope of modals (for an extensive discussion see Condoravdi 2002). Thus, at least the second proposal is not very convincing. Alternatively, Condoravdi (2002) proposed that the meaning of modals combines a modal element: quantification over alternative worlds, with a temporal element: extending the evaluation time for the phrase in the scope of the modal forward. While this way Condoravdi (2002) can account for the future-shift of the evaluation time,

---

<sup>34</sup>In general, the extra universally quantifying modal in the antecedent weakens the truth conditions of the conditional considerably. How convincing this interpretation rule is strongly depends on which modal bases can be associated with the two modals conditionals are predicted to contain.



(where  $i, j$  and  $k$  are indices,  $R_A$  is an accessibility relation the antecedent refers to,  $R_C$  the accessibility relation the consequent refers to,  $TENSE_A$  the tense marked in the antecedent,  $TENSE_C$  the tense marked (on the modal, if present) in the consequent,  $\psi_A$  the property described in the scope of the tense in the antecedent and  $\psi_C$  the property described in the scope of the tense/modal in the consequent. The operation  $*$  closes a modal accessibility relation with respect to the future:  $iR^*j \Leftrightarrow \exists k : iRk \ \& \ t_k = t_j$  where for any index  $l$ ,  $t_l$  denotes the time component of  $l$ . The operation  $*$  is added by the semantics of *IF* to the modal base of the consequent. Summarily, the meaning assigned to conditionals proposed by Kaufmann can be described as follows.)

$$\begin{aligned} & \forall j : [iR_C^*j \ \& \ antecedent(j)] \Rightarrow TENSE_C(\psi_C)(j) \\ = & \forall j : [iR_C^*j \ \& \ \forall k : jR_Ak \Rightarrow TENSE_A(\psi_A)(k)] \Rightarrow TENSE_C(\psi_C)(j) \end{aligned}$$

Figure 6.7: Kaufmann's approach to conditionals

the way she describes the modal meaning of *will* makes it difficult for her to distinguish between the meaning of bare present tense sentence like (125a) and sentences with an explicit modal as in (125b), or, in other words, to account for the observation that statements with *will* come without the certainty condition. Condoravdi (2002) proposes that *will*  $\psi$  holds in case in all ontic alternatives at the evaluation time of the modal  $\psi$  is true. In consequence *will*  $\psi$  means that  $\psi$  is settled at the evaluation time of the modal (which is normally the utterance time). This should rather be the meaning of (125a) than of (125b).

(125) a. The train arrives tomorrow at 9 pm.

b. Peter will arrive tomorrow at 9 pm.

Kaufmann (2005) can also account for the forward shift of the evaluation time in the scope of present tense modals. This is an immediate consequence of the fact that (i) he analyzes the present tense as referring to the present or the future, and (ii) Kaufmann interprets the tense marked on the modal in scope of the modal. He can also account for the observation that a modal statement like (125b) does not come with the certainty condition of (125a). This is explained by the different modal force he assigns to the modal operator described by *will* and to the hidden modal in sentences with a bare present tense, which simply universally quantifies over all accessible worlds. However, Kaufmann's analysis also shows some shortcomings. The introduction of covert modal elements in all tensed sentences is a high price to be paid for the welcome predictions this approach makes. For instance, now it is predicted that also simple, modal free tense sentences should allow for all the different readings possible for modal sentences. Furthermore, because he interprets the tense marked on an overt modal in the scope of the modal, Kaufmann cannot account for modals marked with the simple past and evaluated in the past as in (114).<sup>35</sup> In English, most modals with a syntactic past tense have no longer any uses where they refer to the past. But in earlier days many more modals in English had past interpretation, as have many modals in related languages. Kaufmann has to assume a completely different semantics for them. Furthermore, Kaufmann cannot explain why these modals referring to the past nevertheless extend the evaluation time of the property in their scope to the future (see again (114)). He assumes that the present tense marked on a modal is responsible for that. But in examples like (114) the modal is not marked with the present tense.

So much for approaches to the first, preliminary observation discussed in this section: the future-shift of evaluation times. Now, we come to proposed explanations for the puzzle of the shifted temporal perspective. This was the observation that in conditional and modal contexts the reference time of the interpretation of tenses can also be shifted to the future. One of the few approaches that do discuss this puzzle is Dowty (1982). Dowty focuses on one particular instantiation of the puzzle of the shifted temporal perspective: the observation that a simple past in relative clauses of statements with *will* can obtain a past-in-the-future interpretation. To account for this he proposes a double indexed tense logic. Every expression is interpreted with respect to an index for the utterance time and an index for the evaluation time. The simple past and the simple present shift the evaluation index backward (simple past) or to the utterance time (simple present). The future operator that Dowty (1982) takes to represent the meaning of *will*, however, shifts both the evaluation time as well as the utterance time to some future time. Crouch (1993) discusses an extension of this theory to conditionals.

---

<sup>35</sup>One might argue that the past tense on the modal is a sequence of tense phenomenon. Then, this problem might disappear.

“In the same vein, we could also define the effects of the conditional as follows

$$\llbracket IF(\psi, \phi) \rrbracket^{s,e} = 1 \text{ iff } \llbracket \psi \rrbracket^{e',e'} = 1 \text{ implies } \llbracket \phi \rrbracket^{e',e'} = 1 \text{ for some } e' > s.$$

That is, the conditional acts as though it were within the scope of a FUT operator. This would predict that in

- (126) If I smile when I get out, the interview went well.  
*IF(PRES  $\psi$ , PAST  $\phi$ )*

the antecedent present tense refers to some time in the future, and the consequent past tense refers to some time preceeding it.” (Crouch, 1993: 53)

Crouch criticizes the approach of Dowty (1982) for not being able to account for readings of examples like (124a) where the interpretation of the present tense in a relative clause of a *will* statement is not shifted to the future. Furthermore, he points out that the rule for the interpretation of conditionals is not able to account for the complex temporal relations between antecedent and consequent. For instance, conditionals with present tense in antecedent and consequent are evaluated as referring to the same future time in antecedent and consequent. One problem that we might add is that this rule for conditionals totally ignores the fact that the future-shift of the reference time only occurs for the consequent and not for the antecedent.

A very interesting approach towards explaining the apparent deictic shift in conditional and modal contexts is brought forward by Crouch himself. Fundamental in his approach is the distinction he makes between the time at which some sentence is asserted – the standard utterance time – and the time at which the sentence is verified, i.e. at which the information it contains is indeed updated to the information state. In normal sentences without modals and conditionals both temporal indexes denote the same time. Hence, the update occurs as soon as the assertion is made. Modal and conditional sentences, however, can express that something will be verified in the future. According to Crouch (1993), a sentence like (127a) states that the postman is at the door, but only demands that the statement is verified at some time in the future. That means that in contrast to (127b) at the assertion time the speaker does not have to have direct evidence for his claim.

- (127) a. That will be the postman.

- b. That is the postman.



Furthermore, Crouch proposes that the tenses do not come with one but with two deictic centers, one corresponding to the assertion time and the other to the verification time. Hence, in the context of conditionals or modals where there can be a difference between those two indexes they can refer to the utterance time or to the (future-shifted) verification time.<sup>36</sup> This approach can explain the puzzle of the shifted temporal perspective. But there is a price to be paid for that. The distinction of the level of verification to semantic update adds a lot of complexity to the semantic system. Some linguists also claim that, intuitively, it does not give a correct description of the meaning of sentences like (127a) (see Condoravdi 2003). Crouch (1993) himself admits that if we can do without the additional level of verification, we should dismiss it. In section 6.4 we will show that this is indeed possible.<sup>37</sup>

Also Kaufmann (2005) can account for the shift of the reference time for the interpretation of tenses in the consequent of indicative conditionals. Because in this approach *IF* extends the time of the indices in the modal base of the consequent forward and the tenses are interpreted with respect to these indices (see figure 6.7), their reference point may thus lie in the future as well. The problem is that Kaufmann (2005) – as much as the extension of Dowty (1982) discussed above – predicts a future-shift of the reference time for the antecedent as well: also the evaluation index of the antecedent is set by the elements of the modal base of the consequent, and, thus, shifted forward.

### 6.3.3 Summary

In this section we have discussed two observations that have to be explained by any approach to the meaning of the tenses in the context of conditionals (and modals).

- (i) We have to account for the future-shift of the evaluation time for the interpretation of present tense sentences and phrases in the scope of a modal. Furthermore, we have to explain why future readings of the present tense for simple sentences always come with the certainty condition, while this is not (normally) the case for future uses of the present tense in the antecedent of conditionals and for future evaluations of the phrase in the scope of a modal.
- (ii) We have to account for the puzzle of the shifted temporal perspective, i.e. we have to explain why the reference time for the interpretation of tenses

---

<sup>36</sup>Some minimality constraint ensures that the evaluation time of the consequent of a conditional.

<sup>37</sup>There are other problems concerning Crouch's approach to the semantics of subjunctive conditionals, that we will not discuss in detail here. Crouch (1993) defends a past-as-past approach to subjunctive conditionals. We have criticized this line of approach already in the last section.

can be shifted to the future in the consequent of conditional sentences and in relative clauses in the scope of modals. Furthermore, we have to make sure that the same future shift is not predicted for the antecedent of conditionals and that past tense antecedents cannot lead to a back-shift of the reference time in the consequent.

We discussed a number of proposals made to account for these observations. All of them were found deficient on some points. Nevertheless there is also a lot to learn and to build on in each of them. Particularly, Kaufmann (2005) plays an important role for the theory that will be introduced in the next section. This approach has been used as starting point for the development of the present proposal. One of the main motivations behind the present work was to improve on Kaufmann (2005), in particular, to do without the hidden modal operators this approach assumes and the non-standard treatment of the reference time for the interpretation of the tenses. We will propose a different treatment for the ontic reading that allows us to account for all the observations made here, especially for the puzzle of the shifted temporal perspective.

## 6.4 The proposal

### 6.4.1 An introduction

In this section we are going to propose a compositional approach to the semantic meaning of English conditional sentences. In particular, this approach will provide meanings for the tenses, the perfect and modals occurring in such sentences. We will see that the proposal is able to account for the puzzle of the missing interpretation as well as the puzzle of the shifted temporal perspective. Furthermore, the approach will extend our work from the previous chapter. That means that the compositional theory for conditionals developed here will incorporate the approach developed in Chapter 5.

Before we come to the details of the approach, let us first outline some of the central claims and ideas it builds on. When we discussed the puzzle of the missing interpretation we distinguished between two ways to solve it. First, there is a class of approaches that tries to maintain the standard interpretation of the simple past and the perfect and looks for an explanation of the missing interpretation rather in the logical structure of conditionals and the way the past and the perfect contribute their meanings within this structure. A second class of approaches claims instead that the standard interpretations of the past tense and the perfect are not correct – at least in the context of subjunctive conditionals. In contrast to the first line of approach, proposals adopting this idea can stick to a classical structural analysis of conditionals that stays close to their surface form: *what you see is what you get*.

The approach that will be introduced here follows the second line of approach. Thus, the syntactic structure proposed for conditionals is very close to what you see on the surface; everything is interpreted *in situ*. To account for the puzzle of the missing interpretation we will propose that not everything is interpreted as what it looks. More particularly, we claim that the realization of the semantic simple past in English is the same as the realization of the subjunctive mood. In other words, the syntactic simple past is semantically ambiguous. This claim predicts that semantic effects of the simple past are missing exactly in those cases where the simple past morphology is interpreted as the subjunctive. Actually, we will propose something similar for the past perfect: the past perfect form also allows for two interpretations; a standard interpretation and an interpretation as counterfactual mood. As we will see, this allows us to account for the puzzle of the missing interpretation.

Another distinguishing feature of the approach introduced below is that no reference is made to modal bases, neither to interpret modalities nor for the meaning of conditionals. The function these modal bases fulfill in approaches like Kratzer (1979, 1981), Ippolito (2002), Condoravdi (2002), and Kaufmann (2005) is now fulfilled by two different semantic interpretation functions. The description of these two functions is based on the formalization provided for the two readings of conditionals distinguished in Chapter 5: the epistemic reading and the ontic reading. But now we claim that there are not only two ways to understand *would have* conditionals, but that the ambiguity of the conditionals is based on two fundamentally different ways to interpret language, which stand for two different ways to act with language. The epistemic interpretation function corresponds to a descriptive language use. This interpretation function tells you how to change your information state in case the updated sentence is taken to provide new information about the actual world. The way it is defined here more or less agrees with dynamic interpretation as we know it – particularly from the Amsterdam school of dynamic semantics. The ontic reading is based on a prescriptive language use. It turns every world considered possible into one where the sentence that is interpreted is true.

The systematic distinction of different interpretation functions on the level of all expression is, even though not unique, quite unusual in formal semantics. It appears to go against the central goal of classical formal semantics, which is to remove all the ambiguities in natural language. But we do not stand in opposition to this tradition, because we propose that these two interpretation functions represent two different speech act types. It is, thus, the way a language distinguishes between different speech acts that resolves the potential ambiguity the approach produces.

In the context of this work we are only interested in assertions, or, in other words, the descriptive use of language. However, we will propose that there are some lexical items whose descriptive interpretation makes reference to the ontic

interpretation function. Among these are the modals *will*, *would*, *may* and *might*, as well as the sentence connective *if*. Only for these items will the distinction of two interpretation functions in the present framework indeed predict an ambiguity, because in their epistemic update they can make reference to the ontic interpretation function as well as the epistemic one.

The semantic theory that will be developed here is a type-theoretic version of compositional dynamic semantics. *Type-theoretic* means that the expressions of the formal language this semantics interprets are assigned types that restrict how expressions of the language can be defined as well as what meanings can be assigned to the expressions. *Compositional* means that we will assume that the meaning of complex expressions is determined by the meaning of the parts they consist of and the way they combine. In consequence, the meaning of complex expressions can be defined inductively by specifying (i) the translation of basic expressions, and (ii) how the translation of a complex expression depends on its parts. Finally, the theory is a *dynamic* theory of meaning, because it takes the meaning of a sentence to be not its truth conditions but its context change potential, i.e. how some epistemic state is transformed by updating it with the sentence.

Four ingredients have to be provided in order to describe such a theory.

- (A) We have to describe the formal language for which a semantics is provided and link this language to English.
- (B) We have to define the class of models with respect to which the language is interpreted.
- (C) We have to lay down the rules of interpretation for basic expressions.
- (D) Finally, we have to say how the interpretation of complex expressions depends on the interpretation of its parts.

The remainder of the section is structured around this list of requirements. Thus, we will start by introducing the language and the model. Then the heart of the theory, the interpretation rules for all basic expressions, will be described and motivated. There is no need for an independent section on part (D) of this scheme. We assume only one rule for how the meaning of a complex expression can be calculated from the meaning of its parts. This is the rule of functional application.

### 6.4.2 The language

In this section we are going to define the formal language  $\mathcal{L}$  that we take to mirror English (conditional) sentences in the aspects relevant for our analysis. We will

provide a standard definition of a type-theoretic formal language. In a second step we will additionally provide a rough description of the logical form for a fragment of English, that further restricts the admissible sentences of  $\mathcal{L}$ . This description, together with the type-theoretic restrictions on well-formedness, will define the well-formed *sentences* of the language  $\mathcal{L}$ .

We start the description of the formal language  $\mathcal{L}$  by defining the types an expression in  $\mathcal{L}$  can have. Possible types are defined in an recursive manner, providing a set of basic types and the way they can be combined. For this language we distinguish four basic types:  $i$  for times,  $s$  for states of affairs (these will be partial interpretation functions for the set of proposition letters of  $\mathcal{L}$ ),  $n$  for natural numbers (used as indexes for subordinate contexts), and  $t$  for truth values.

**6.4.1. DEFINITION.** (The set of types)

The set of types  $\mathcal{T}$  is the smallest set such that

- (i)  $i, s, n, t \in \mathcal{T}$ ,
- (ii) if  $a, b \in \mathcal{T}$  then  $\langle a, b \rangle \in \mathcal{T}$ .

Because the semantics that will be assigned to the language  $\mathcal{L}$  is a dynamic semantics, formulas of  $\mathcal{L}$  are not interpreted as functions from states of affairs to truth value (type  $\langle s, t \rangle$ ). Instead, they denote functions from cognitive states to cognitive states. A cognitive state is a partial assignment of sets of states of affairs to natural numbers - we will call these numbers *indexes*. A precise definition is given below. Thus, the type of a cognitive state is  $\langle n, \langle s, t \rangle \rangle$  and the type of a formula, in consequence, is  $\langle \langle n, \langle s, t \rangle \rangle, \langle n, \langle s, t \rangle \rangle \rangle$ . To improve readability, we will abbreviate the notation of the types a bit and write  $[\alpha_1 \dots \alpha_n]$  for  $\langle \alpha_1, \langle \alpha_2, \langle \dots \langle \alpha_n, \langle \langle n, \langle s, t \rangle \rangle, \langle n, \langle s, t \rangle \rangle \rangle \dots \rangle \rangle \rangle$ . With this notational convention the type of formulas becomes, for instance,  $[]$ . Next, the vocabulary of  $\mathcal{L}$  is defined. The vocabulary is a set of basic expressions of  $\mathcal{L}$  plus their type.

**6.4.2. DEFINITION.** (The vocabulary of  $\mathcal{L}$ )

The vocabulary of the type-theoretical language  $\mathcal{L}$  for the set of types  $\mathcal{T}$  contains:

- (i) for the type  $i$  an infinite set of variables  $VAR_i$ ,
- (ii) for the type  $[i]$  a finite set of constants  $\mathcal{P}$ ,
- (iii) operators  $\wedge, \vee, IF$  of type  $[] []$ ,
- (iv) an operator  $\neg$  of type  $[]$ ,
- (v) the brackets ( and ),
- (vi) an enumerable set of operators  $PAST_n, PRES_n$  of type  $[[i]]$ ,
- (vii) an enumerable set of operators  $WOLL_n, MOLL_n$  of type  $[[i]i]$ , and
- (viii) an enumerable set of operators  $PERF_n$  of type  $[[i]i]$ ,
- (ix) operators  $IND, SUBJ$  and  $COUNT$  of type  $[]$ .

The types restrict the way basic expression of  $\mathcal{L}$  can be combined in more complex expressions. The next definition describes which combinations are possible.

**6.4.3. DEFINITION.** (The expressions of  $\mathcal{L}$ )

- (i) If  $a$  is a variable or a constant of type  $\alpha$  in  $\mathcal{L}$ , then  $a$  is an expression of type  $\alpha$  in  $\mathcal{L}$ .
- (ii) If  $a$  is an expression of type  $\langle\alpha, \beta\rangle$  in  $\mathcal{L}$ , and  $b$  is an expression of type  $\alpha$ , then  $(a(b))$  is an expression of type  $\beta$  in  $\mathcal{L}$ .
- (iii) Every expression in  $\mathcal{L}$  is to be constructed by means of (i) - (ii) in a finite number of steps.

Brackets around expressions will be left out if this will not cause ambiguity. Finally, we define what a formula of  $\mathcal{L}$  is.

**6.4.4. DEFINITION.** (The formulas of  $\mathcal{L}$ )

The formulas of  $\mathcal{L}$  are the expressions of type  $[]$ .

We will now provide additional restrictions on what well-formed sentences of our formal language are. The difference between these restriction and what has been described above can be best understood as follows. Type theoretical restrictions are driven by semantics. The type of a basic expression determines the type of the function that can serve as meaning for this expression. The only semantic rule of composition of meanings that we will allow is functional application. Therefore, the semantics predicts that an expression  $A$  can only combine with an expression  $B$ , if the function that represents the meaning of one of the expressions can be an argument of the function that represents the meaning of the other. This is all that the definition of a well-formed expression says. This allows for many structures that do not correspond to well-formed English sentences. For instance, the modal operators and the perfect operator can be iterated arbitrarily often and mood is not obligatory for formulas in  $\mathcal{L}$ .<sup>38</sup> In principle, there is nothing wrong with such a result. Some semantically well-formed expressions may not be syntactically well-formed or are never used for other reasons. We need additional, syntactical restrictions on the language  $\mathcal{L}$ , because the problems we want to solve are about an apparent misfit between form and meaning of conditional sentences and we are going to propose that at least part of the solution lies in the logical form assigned to English conditionals.

We will give only a rough outline of the logical form of the fragment of English we are interested in. The way we describe LF here surely simplifies and ignores many aspects of syntax, but a correct description of English syntax is not the primary topic of concern for the present investigation. This part of the theory may be elaborated in future work. We distinguish two classes of syntactic categories. First, there are the *lexical* categories PROPERTY, MODAL, ASPECT, TENSE, MOOD, CONNECTIVE. Each of them relates to a set of expressions of the vocabulary of  $\mathcal{L}$ . Second, a set of *complex* categories is needed that contains

---

<sup>38</sup>Maybe it is possible to get rid of some of these unwanted structures again by semantic constraints, but the semantics we provide here do not provide these constraints.

the following members: sentence phrases SP, tense phrases TP, modal phrases MP, aspectual phrases AP, and inflectional phrases IP. A set of rules tells us how complex categories can be decomposed into one or more other categories. We will not specify these rules here one by one, but just illustrate their output. Ignoring connectives, the syntactic structure of English sentences at the level of LF is assumed to appear as given in figure 6.8. Because English will not be analyzed to the level of predicate structure, we do not distinguish between nominal phrases and verbal phrases. The syntactic analysis stops at the level of IPs. The primitives, called properties, are roughly phrases consisting of the main verb plus its arguments.

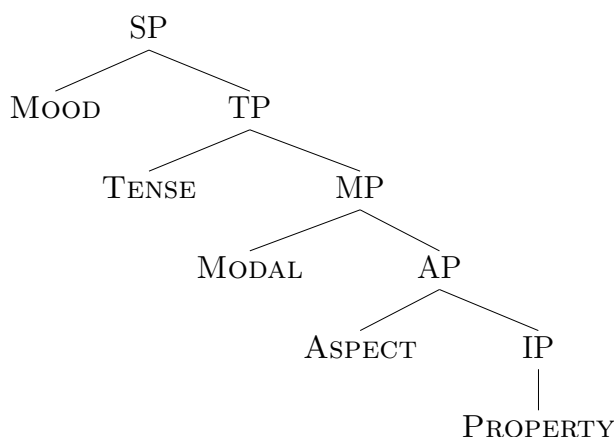


Figure 6.8: The syntactic structure at the level of LF

The logical form is additionally specified by the lexicon. The lexicon lists for every lexical category a number of lexical entries. Every lexical entry combines a meaning with a set of syntactic features and a realization of this combination of meaning and syntactic features. Not in every case all three variables, meaning, set of syntactic features, and realization, have to be specified. Sometimes the set of syntactic features is empty. There are also combinations of meaning and syntactic features that have no realization. We can even have forms that only have a syntactic function, but do not contribute by themselves to the semantic meaning of a sentence. In figure 6.9 on page 206 we sketch the lexicon that connects the fragment of English considered here with syntax and semantics. Figure 6.9 will not provide the semantic entries for the basic expressions. They are described in the next section.

The syntactic features of entries in the lexicon impose additional restrictions on the syntax of the logical form of an English sentence. For every syntactic feature there exists a positive and a negative version. Features move up in the tree. If at one node the negative and the positive version of some feature meet, both are erased from the list of features. If two different lexical entries contribute

the same feature, then both features independently move up in the tree and have to find a counterpart. The following syntactic features are distinguished: for the category MOOD *ind*, *sub*, and *count*, for the category TENSE *pres* and *past*, and for the category ASPECT *perf*. We define a *sentence* of  $\mathcal{L}$  as a formula of  $\mathcal{L}$  that is generated by a tree as described above where the set of features in every SP node is empty.

We assume that the morphological category of the simple past is ambiguous and expresses two different syntactic feature combinations: either it asks for the past tense operator *PAST* or for the mood operator *SUBJ*. If the simple past is interpreted as mood feature, then the verb also carries a  $[-pres]$  feature. Hence, the subjunctive obligatory combines with the present tense. A similar ambiguity is also proposed for the syntactic perfect. The auxiliary *have* is either interpreted as the perfect operator or selects for the counterfactual mood. In the second case it does not carry a tense feature like the simple past. The counterfactual mood is only realized if some other past tense marking in the sentence asks for the subjunctive mood. In this case, as we have just explained, this past tense marking also demands the present tense operator *PRES*.

The lexicon distinguishes two modals: *WOLL* realized, depending on the features, as *will* or *would*, and *MOLL* realized as *may* or *might*. The choice of the realization depends on the pre- or absence of the morphological simple past. This marker can be either interpreted as the feature  $[-past]$  or as the feature combination  $[-subj, -pres]$ . However, for *MOLL* the feature set  $[-past]$  has no lexical entry. It is, thus, predicted that *might* cannot be interpreted as referring to the past.<sup>39</sup>

---

<sup>39</sup>It has often been noticed in the literature that there are no uses of ‘*might*’ in standard American or British English that evaluate the modal in the past. Stowell (2002), however, claims that there are some local varieties of English where these uses still exist.



LEXICONCategory: PROPERTY

semantic expression	type	syntactic features	realization
$P$	[i]	[-ind, -pres]	<i>Mary-drinks-all-the-wine</i>
$P$	[i]	[-ind, -past]	<i>Mary-drank-all-the-wine</i>
$P$	[i]	[-subj, -pres]	<i>Mary-drunk-all-the-wine</i>
$P$	[i]	[-perf]	<i>Mary-drunk-all-the-wine</i>
$P$	[i]	[]	<i>Mary-drunk-all-the-wine</i>
$P$	[i]	[]	<i>Mary-drink-all-the-wine</i>
...	...	...	...

Category: MODAL

semantic expression	type	syntactic features	realization
$WOLL_n$	[[i]i]	[-ind, -pres]	<i>will</i>
$WOLL_n$	[[i]i]	[-ind, -past]	<i>would</i>
$WOLL_n$	[[i]i]	[-subj, -pres]	<i>would</i>
$MOLL_n$	[[i]i]	[-ind, -pres]	<i>may</i>
$MOLL_n$	[[i]i]	[-subj, -pres]	<i>might</i>

Category: ASPECT

semantic expression	type	syntactic features	realization
$PERF_n$	[[i]i]	[-ind, -pres, +perf]	<i>have</i>
$PERF_n$	[[i]i]	[+perf]	<i>have</i>
	[[i]i]	[-count, -pres]	<i>have</i>
$PERF_n$	[[i]i]	[-ind, -past, +perf]	<i>had</i>
$PERF_n$	[[i]i]	[-subj, -pres, +perf]	<i>had</i>
	[[i]i]	[-count, -subj, -pres]	<i>had</i>

Category: TENSE

semantic expression	type	syntactic features	realization
$PRES_n$	[[i]]	[+pres]	*
$PAST_n$	[[i]]	[+past]	*

Category: MOOD

semantic expression	type	syntactic features	realization
$IND$	[[[]]]	[+ind]	*
$SUBJ$	[[[]]]	[+subj]	*
$COUNT$	[[[]]]	[+count, +subj]	*

Category: CONNECTIVES

semantic expression	type	syntactic features	realization
$IF$	[[[] []]]	[]	<i>if</i>
$\neg$	[[[]]]	[]	<i>not</i>
$\wedge$	[[[] []]]	[]	<i>and</i>
$\vee$	[[[] []]]	[]	<i>or</i>

Figure 6.9: The lexicon of our fragment of English

### 6.4.3 The model

In this section we will describe the model with respect to which the formal language  $\mathcal{L}$  just introduced will be interpreted. We start by providing the general definition of a model, which is quite a complex structure. Then, step by step, we will explain how all the elements of a model have to be interpreted.

#### 6.4.5. DEFINITION. (Model)

A *model* for the language  $\mathcal{L}$  is a quintuple  $M = \langle S, \langle T, < \rangle, \langle C, U \rangle, I, now \rangle$  where  $S$  is a set of states of affairs,  $\langle T, < \rangle$  a time structure,  $\langle C, U \rangle$  a law structure,  $I$  a function mapping states of affairs to partial interpretation functions for property letters ( $I : S \longrightarrow (\mathcal{P} \times T \longrightarrow \{0, 1\})$ ), and  $now$  a function associating every state of affairs with a specific time ( $now : S \longrightarrow T$ ).

This model contains two sets of basic entities: a set of states of affairs  $S$  representing roughly alternative ways the actual world might be, and a set  $T$  of time points. We make very few restrictions on how time is structured. We only demand that times are ordered by a partial order<sup>40</sup>.

#### 6.4.6. DEFINITION. (Time)

A *time structure* is a tuple  $\langle T, \leq \rangle$  where  $T$  is a set of times and  $<$  a strict partial order on  $T$ . An *interval*  $I$  is a subset of  $T$  such that  $\forall t_1, t_2, t_3 \in T : (t_1, t_2 \in I \ \& \ t_1 \leq t_3 \leq t_2) \Rightarrow t_3 \in I$ .  $I(T)$  is the set of all intervals of  $T$ . We extend the definition of the order  $<$  to intervals as follows:  $\forall I_1, I_2 \in I(T) : I_1 < I_2 \Leftrightarrow \forall t_1 \in I_1 \forall t_2 \in I_2 : t_1 < t_2$ .

The definition of a model we adopt assumes a fixed time structure for all states of affairs. Alternatively, we could have assigned to each state of affairs its own time structure. The reason why we have chosen the uniform structure is mainly that it simplifies matters strongly. Among others, van Benthem (1983) has argued against such a simplification. His point is that while proper names like *Nixon* can be taken unproblematically to denote the same individual in alternative worlds, things are much less clear for temporal designators. The problem might also show up on the technical side, when one introduces events into the model. Many authors have argued that there has to be a strong relation between the event structure and the temporal structure of possible worlds (see, for instance, Kamp & Reyle 1993). Some of them have even proposed that the time structure is a product of the event structure. But because worlds certainly differ in the events taking place in them, as a consequence, they may also differ in the times they distinguish. However, because we will not make the step to event semantics here, this problem does not play a role in our considerations. This is not to say that there is in general no need to give up the simplification to one uniform temporal

---

<sup>40</sup>A strict partial order  $<$  is a binary relation of a domain  $D$  that is irreflexive and transitive.

structure in future work.

In the previous chapter, we have seen that laws play a central role in the interpretation of *would have* conditionals. The formalization of both meanings we distinguished in Chapter 5 for *would have* conditionals made reference to a set of laws. Because the present chapter builds on the work of the previous one, we again need to represent laws within a model. As in the last chapter, we make the simplifying assumptions (i) that the set of relevant laws for a cognitive state is clearly defined, hence, there is no uncertainty about which laws are taken to hold, and (ii) that the update with a sentence  $\phi$  cannot change this set of laws. We will also adopt the way laws were represented in Chapter 5. Hence we will store causal laws and analytical/logical laws separately, the first using a causal structure, the second with a set of histories. Below, the definition of a *law structure* is given, which is very similar to the notion of a model introduced in definition 5.6.13 of section 5.6.3.1. The definition of a causal structure and what it means for a causal structure to be rooted are directly adopted from this section (definition 5.6.10 on page 141 and definition 5.6.11 on page 141) and will not be repeated here.

#### 6.4.7. DEFINITION. (Law structure)

A *law structure* is a tuple  $L = \langle C, U \rangle$ , where  $C$  is a causal structure and  $U$  is a set of complete interpretation functions  $u : \mathcal{P} \times T \longrightarrow \{0, 1\}$ .

The remaining two parameters  $I$  and *now* of a model  $M$  characterize the states of affairs in  $S$  of  $M$ . The function  $I$  assigns to every state of affairs an interpretation function for the property letters  $P \in \mathcal{P}$  at times  $T$ . An important property of the function  $I$  is that these interpretation functions may be partial. Hence, a state of affairs  $s$  may not decide for all times  $t$  and all property letters  $P$  whether  $P$  is true at  $t$ . To simplify notation we will use  $w_s$  to refer to  $I(s)$ . The set  $\text{dom}(s)$  denotes the set of tuples  $\langle P, t \rangle$  of elements of  $\mathcal{P}$  and  $T$  for which  $w_s$  is defined. Because possible worlds are normally complete interpretation function, we will not refer to  $w_s$  or  $I(s)$  as worlds, but call them simply (partial) interpretation functions. The term *worlds* will only be used when reference is made to complete functions  $f : \mathcal{P} \times T \longrightarrow \{0, 1\}$ . For instance, the elements of the set  $U$  of a law structure  $\langle C, U \rangle$  are worlds in this sense. Finally, the function *now* assigns a temporal perspective to a state of affairs. This temporal perspective will play an important role for the interpretation of the tenses: it will set their reference time. Conceptually, this time will be interpreted as the actual time; the temporal deictic center of a state. Thus, according to this approach every state of affairs does not only have its own opinion on what the actual world is, but also on what the actual time is.<sup>41</sup> We will use  $t_s$  to refer to  $\text{now}(s)$ .

---

<sup>41</sup>Because of the restrictions of the formal language chosen here, this is the only parameter of the utterance context needed for interpretation. For a more extended language one might need as well a parameter for the speaker, the hearer, the location of the conversation, etc..

There is another parameter of a state of affairs that is needed for interpretation, but this parameter is not set by the model. For the interpretation of the variables for times in the language we need a function assigning referents to these variables.

#### 6.4.8. DEFINITION. Assignment

An *assignment* (or *assignment function*) is a function  $g : S \longrightarrow (VAR_i \longrightarrow T)$  that assigns to every element of  $S$  a function that interprets the temporal variables.

We will use  $g_s$  to refer to the function  $g$  assigns to some state of affairs  $s$ . Also  $g_s$  will be called an assignment or an assignment function. This should not lead to any confusions. The assignment function  $g_s = g(s)$  of a state of affairs encodes the dynamic information of temporal referents that already have been introduced in the discourse. It will play a role in the interpretation of the tenses, as well as the perfect and the modals.

For our later considerations it is very important that there are sufficient states of affairs in the model. We will simply assume that this is the case, that means that for every possible value of the three parameters  $w_s$ ,  $t_s$ , and  $g_s$  there is some  $s$  in  $S$  that takes these values.

As said above, we will describe the meaning of the language  $\mathcal{L}$  in terms of dynamic semantics. Formulas of  $\mathcal{L}$  do not receive a truth value with respect to a (state of affairs in a) model, as in classical static semantics, but they will denote functions from cognitive states to cognitive states. The structure of cognitive states is what interests us next. It will be defined stepwise. First, we select a subclass of states of affairs that gives a convincing description of epistemic alternatives an agents in a discourse might consider possible. We will call this subclass *possibilities*.<sup>42</sup>

#### 6.4.9. DEFINITION. (Possibilities)

Let  $M = \langle S, \langle T, < \rangle, \langle C, U \rangle, I, now \rangle$  be a model for the language  $\mathcal{L}$  and  $g$  an assignment function.  $S_U \subseteq S$  is the set of all states of affairs  $s \in S$  for which the following holds:  $\exists u \in U : s \subseteq u$ . A *possibility* with respect to  $M$  and  $g$  is a state of affairs  $p \in S_U$ , where  $w_p$  restricted to  $\{t' \mid t' \leq t_p\}$  is a complete function and  $g_p(d_0) = t_p$ .  $W_{M,g}$  is the set of all possibilities with respect to  $M$  and  $g$ .  $p[x/t']$  is the possibility  $p'$  that is like  $p$  except that  $g_{p'}$  is defined for  $x$  and  $g_{p'}(x) = t'$ .

A possibility is thus a state of affairs distinguished by three properties: (i) the (partial) interpretation function assigned to it by  $I$  is complete for all times before or equal to its temporal perspective  $t_p$ ,<sup>43</sup> (ii) the variable  $d_0$  is assigned the value

<sup>42</sup>The definition of the notion *possibilities* given here will be adapted at a later point in this chapter.

<sup>43</sup>Thus, only the future may be undefined for certain combinations of properties and times.

of the temporal perspective, and (iii)  $p$  does not violate any analytical/logical laws encoded in the law structure  $\langle C, U \rangle$  of the model.

Next, we define the notion of a *basic state*. A basic state is a set of possibilities. It represents a possible epistemic state of some agent.

**6.4.10. DEFINITION.** (Basic state)

Let  $M = \langle S, \langle T, < \rangle, \langle C, U \rangle, I, now \rangle$  be a model for the language  $\mathcal{L}$  and  $g$  an assignment function. A *basic state*  $R$  is a set of possibilities with respect to  $M$  and  $g$ .

We do not demand that the possibilities in a basic state  $R$  all assume the same actual time, that is, for all  $p, p' \in R : t_p = t_{p'}$ . The reason is that in this case the agent holding the beliefs represented by a basic state would know what the actual time is, which is not intuitively warranted.

A cognitive state represents the information state of an agent (or the common information state of a group of agents) involved in a conversation. It will be formalized by a partial function assigning basic states to natural numbers. These numbers do not stand for different agents, but for different (subordinated) contexts distinguished by one agent. The value such a function assigns to 0 is the standard discourse context we are familiar with from dynamic semantics. It represents the information about the actual world available in a cognitive state. But for the semantics of the fragment of English we are considering, we cannot do just with just one basic state. There are expressions in this fragment that involve the consideration of hypothetical, subordinated states. This is the case for conditionals, but also the modalities.<sup>44</sup>

**6.4.11. DEFINITION.** (Cognitive state)

Let  $M = \langle S, \langle T, < \rangle, \langle C, U \rangle, I, now \rangle$  be a model for the language  $\mathcal{L}$  and  $g$  an assignment function. A *cognitive state* is a partial function  $c : \mathbb{N} \rightarrow \wp(S)$  with the following two properties: (i)  $c$  is defined for 0, and (ii) there exists some  $n \in \mathbb{N}$  such that  $c$  is defined for all natural numbers smaller than and equal to  $n$  and undefined for all natural numbers bigger than  $n$ . For every cognitive state  $c$ , we let  $\eta(c)$  denote this natural number  $n$ .  $\perp$  is the set of *absurd cognitive states*, i.e. cognitive states  $c$  such that  $\exists i \in \mathbb{N} : c(i) = \emptyset$ . For  $i \in \mathbb{N}$ ,  $c_i$  refers to the value of  $c$  at index  $i$ .

Let  $c$  be a cognitive state,  $i$  a natural number, and  $R$  a basic state.  $c[i/R]$  denotes the cognitive state  $c'$  that is defined like  $c$ , except that  $c'(i) = R$ , if this cognitive state exists.  $c[\eta(c)]$  abbreviates  $c[(\eta(c) + 1)/c_{\eta(c)}]$ .

English sentences do not always contribute information about  $c_0$ , the basic state where all information about the actual world is gathered. They may as well

---

<sup>44</sup>Later on, in the discussion section, we will question whether indeed the introduction of subordinated hypothetical contexts is part of the semantic meaning of these expressions.

provide information about a subordinate, hypothetical context. This phenomenon is known as *modal subordination*. There are certain linguistic mechanisms that govern to which hypothetical context a sentence or a formula refers. This dissertation is not about how these mechanisms work – at least on the sentence level.<sup>45</sup> However, we do make very explicit predictions about the introduction of hypothetical, subordinate contexts. Furthermore, when we describe the semantic meaning of operators like *IF* we make claims about how reference to hypothetical contexts works within sentence boundaries. More particularly, we will assume that the reference context for subsequent updates within one sentence is the basic state assigned to the last number the relevant cognitive state is defined for. This proposal is certainly not correct when applied to how the reference context is determined for independent sentences. It would predict that a new sentence is always about the (hypothetical) context introduced last. To make clear that there is a difference between inter-sentential and intra-sentential update, and also in order to be able to discuss simple cases of inter-sentential modal subordination, we will assume the following simplifying mechanism to determine the reference context for independent sentences:

Algorithm to determine the reference context of independent sentences

The update of a cognitive state  $c$  with sentence  $\psi$  is defined as follows:

- (i)  $c'[\phi]$  where  $c'$  is only defined for 0 and  $c'(0) = c(0)$ , if  $c'[\phi] \notin \perp$ ,
- (ii)  $c[\phi]$ , otherwise.

According to this approach any formula is by default updated to the common ground, the basic state that gathers the information about the actual world, and all previously introduced subordinated contexts are lost. If this update leads to an absurd cognitive state, then the formula is updated to the subordinate basic state introduced last. If again the update leads to an absurd cognitive state, then that's the result. This is, of course, a simplified view on modal subordination. There are examples where reference is made to contexts other than the common ground or the last introduced hypothetical context. There are also counter-examples to the prediction that a subordinate states are lost when reference has been made to a basic state superordinating it. However, because inter-sentential modal subordination is not one of the central issues of this research, we will accept this simplification.

#### 6.4.4 The interpretation of the vocabulary of $\mathcal{L}$

We now come to the central part of the proposal. In the following meanings will be assigned to all elements of the vocabulary of  $\mathcal{L}$ . This together with the rule of composition of meanings, which we introduced earlier, allows us to calculate

---

<sup>45</sup>See Asher & McCreedy (2007) for an approach to conditionals that takes modal subordination very serious

the meaning of arbitrary complex expressions of  $\mathcal{L}$ . We will start with describing the meaning of property letters in  $\mathcal{L}$  and then work through the entire list of operators of  $\mathcal{L}$ . At the end stands the definition of the connective *IF*.

The meanings that are assigned to expressions have to be in accordance with the type of the expression. To be more precise, let us first introduce the domain of interpretation of a type.

**6.4.12. DEFINITION.** (Domain of interpretation)

Let  $M$  be a model for  $\mathcal{L}$ .

- (i)  $D_{s,M} = S$ ,
- (ii)  $D_{i,M} = T$ ,
- (iii)  $D_{n,M} = \mathbb{N}$ ,
- (iv)  $D_{t,M} = \{0, 1\}$ ,
- (v)  $D_{\langle\alpha,\beta\rangle,M} = D_{b,M}^{D_{\alpha,M}}$ .

We now can restate the restriction on the interpretation of an expression of  $\mathcal{L}$  as the claim that the interpretation of an expression of type  $\alpha$  has to be an object in the domain  $D_{\alpha,M}$  of its type.

As explained in the introduction, in the semantics proposed here two interpretation functions will be distinguished: an *epistemic* interpretation function and an *ontic* interpretation function. Both obey the semantic restrictions imposed by the type assigned to an expression. The epistemic interpretation function, corresponding to a descriptive language use, will be captured by the function *Learn*. The ontic interpretation function, based on a prescriptive language use, will be described by the function *Intervene*. Because we restrict our attention to the assertive use of sentences in  $\mathcal{L}$ , on the level of sentences always the interpretation function *Learn* is applied. But as said above, for some elements of the vocabulary of  $\mathcal{L}$  we will propose that their interpretation rule makes reference to the ontic update function *Intervene*. Therefore, we also have to provide the ontic interpretation rules for all expressions of  $\mathcal{L}$ . As we will see, in many cases there is no difference in the interpretation both functions assign to some element of the vocabulary of  $\mathcal{L}$ . For brevity, in these cases we will use  $c[\psi]$  to refer to both,  $\text{Learn}(c, \psi)$  and  $\text{Intervene}(c, \psi)$ . Within one equation each occurrence of  $[\cdot]$  has to refer to the same interpretation function.

We have also mentioned in the introduction that the definition of the functions *Learn* and *Intervene* given here is based on the formalization of the two readings of conditionals provided in the previous chapter. As the reader might recall from Chapter 1 and Chapter 4, this was the original motivation behind the work presented in Chapter 5: to develop a basis on which we can then build a compositional, time-sensitive approach to the meaning of English conditionals. In principle, the only thing we have to do now is adapt the approach developed in Chapter 5 to our new time-sensible model and apply it to our extended

formal language. There is one very common assumption we will make in the present section. As said before, we are concerned here with the interpretation of assertions. An update with an assertion is only successful, if the information the sentence conveys is consistent with the information already encoded by the cognitive state to which the sentence is updated (Stalnaker, 1978). This assertion condition restricts the application of the interpretation functions *Learn* and *Intervene* in many cases to consistent updates. In this case, all the techniques we introduced in Chapter 5 to deal with counterfactual updates are superfluous. The only situation in our framework where this assertion condition is not in force is the update with the antecedent of some conditional. Therefore, we will first introduce simplified versions of the two interpretation functions *Intervene* and *Learn* that build the well-formedness condition of assertions into the semantic update. These are the functions *ALearn* and *AIntervene*.<sup>46</sup> Only when providing the interpretation rule for the conditional connective *IF* will we introduce the full-blooded versions that come without the assertion condition and can deal with inconsistent updates. These later definitions are indeed substantially similar to those proposed in Chapter 5. The conditional connective will then explicitly refer to the extended versions of the interpretation functions *Learn* and *Intervene*.

The distinction of a simplified version of the interpretation functions that build originally pragmatic conditions into semantics is from a conceptual point of view not optimal. Nevertheless, we decided to make this distinction, because it simplifies matters considerably and, thereby, strongly improve the readability of the proposal. In fact, this reduction of interpretation functions to consistent updates is a common practice in semantics. In section 6.4.4.9 we will also provide a more technical motivation for why at least a distinction between *Learn* and *ALearn* is necessary in our framework.

#### 6.4.4.1 The epistemic update with atomic formulas.

The update function *ALearn* defines the update effect a sentence  $\psi$  has on a cognitive state if it is taken to convey new information about the world. It breaks down and delivers an absurd cognitive state in case the sentence is not true in any possibility of the basic state that is updated with  $\psi$ . This is the standard update we are familiar with from dynamic semantics.

When you work with partially defined interpretation functions, it is important to think right from the beginning about how to define the meaning of negation. Intuitively, an update with a negated sentence  $\neg\psi$  should return those possibilities where  $\psi$  is false. Because we work with partially defined interpretation functions, we cannot adopt the standard semantics for negation to produce this result. That means that we cannot define the meaning of negation as some complement of the update with  $\psi$ .<sup>47</sup> The set-theoretical complement of the set of possibilities

<sup>46</sup>The *A* at the beginning of the two names symbolizes the addition of the assertion condition.

<sup>47</sup>The exact definition of the meaning of negation can differ in various systems of dynamic



where some atomic formula  $P(d)$  is true does not only contain those possibilities where  $P(d)$  is false, but also those where the truth-value of  $P(d)$  is not (yet) determined. To deal with this problem, we will adopt a standard solution and define two different epistemic update functions: a positive and a negative version. Both functions have to be specified for every expression.<sup>48</sup> Then, the meaning of negation will be defined as the negative update with the formula in scope of the negation. The crucial difference between the positive and the negative update function is set on the level of an update with atomic formulas. The positive update with an atomic formula  $P(d)$  will select the possibilities where the formula is true, the negative those where the formula is false.

**6.4.13. DEFINITION.** (The epistemic interpretation rule for atomic formulas)

Let  $M$  be a model,  $g$  an assignment function, and  $c$  a cognitive state. For  $P \in \mathcal{P}$  and  $d \in \text{VAR}_i$ , we define:

$$\begin{aligned} \text{ALearn}_{M,g}^+(c, P(d)) &= c[\eta(c)/\{p \in c_{\eta(c)} \mid w_p(P, g_p(d)) = 1\}], \\ \text{ALearn}_{M,g}^-(c, P(d)) &= c[\eta(c)/\{p \in c_{\eta(c)} \mid w_p(P, g_p(d)) = 0\}]. \end{aligned}$$

#### 6.4.4.2 The ontic update with atomic formulas.

The ontic update function *makes* an atomic formula true in a possibility by changing the interpretation function associated with the possibility. It will operate particularly in that space left unexploited by the epistemic update: the regions where the interpretation function is undefined. If the truth value of an atomic formula  $P(d)$  is defined, *AIntervene* behaves exactly like the epistemic update. If the value is undefined with respect to a possibility  $p$ , then *AIntervene* will set this value.<sup>49</sup>

Again, to be able to deal with negation, a positive and a negative version of *AIntervene* have to be distinguished. After update with the positive variant the truth value of an atomic formula is set to 1, if it was not already defined to be 0. In the latter case the possibility is eliminated.<sup>50</sup> After a negative update with *AIntervene* the truth value of  $P(d)$  is set to 0, if it was not defined to be 1. Again, in the later case, the possibility is eliminated.

In order to find a formalization of this general description of *AIntervene* we have to do some thinking. There are no standard approaches we can use like for the epistemic update function, nor can we unreservedly apply the formalization of

---

semantics. One definition that is often used will be introduced later.

<sup>48</sup>Because we have to distinguish positive and negative versions for both interpretation functions, *ALearn*, and *AIntervene*, we work in total with four different update functions for  $\mathcal{L}$ .

<sup>49</sup>One can understand the function *AIntervene* as an extension of the standard update functions that in addition to the assignment parameter can also change the world parameter of a possibility.

<sup>50</sup>This is the effect of the inbuilt assertion condition.

the ontic reading proposed in Chapter 5. The description of *Intervene* provided there was primarily made to deal with the revision case, which is now excluded by the assertion condition. The interesting case we have to deal with now, the situation where a possibility is undefined for the truth value of the atomic formula it is updated with, did not play any role in Chapter 5. In this chapter we did not work with partially defined worlds. Thus, we have to consider the problem how to formalize the function *AIntervene* from scratch.

Assume that we want to apply the positive ontic update function *AIntervene*<sup>+</sup> to a possibility  $p$  and an atomic formula  $P(d)$ . A first idea of how to formalize that *AIntervene*<sup>+</sup> makes  $P(d)$  true in possibility  $p$  might be this: if  $w_p(P, g_p(d))$  is undefined, then the output of *AIntervene* is a possibility  $p'$  that equals  $p$  in every respect except that  $p'$  is defined for  $P$  and  $d$  and  $w_{p'}(P, g(d)) = 1$ . But what we would formalize this way is rather changing  $p$  into a possibility where it is predetermined (at  $t_p$ , the temporal center of the possibility  $p$ ) that  $P$  will be true at  $d$ . This is not what we want for the applications of the ontic interpretation function we are interested in. Remember that we need the ontic update in particular to formalize the interpretation of the antecedent of conditional sentences. Normally, the antecedent of indicative conditionals does not select those possibilities where its truth is predetermined, but where the antecedent just turns out to be true at its evaluation time (recall the discussion in section 6.3.1). Furthermore, the notion of predetermination that underlies the update rule outlined above is not adequate. Predetermination does not just mean that the truth value of some future fact is already set at the utterance time. We have very specific ideas about when some fact about the future can be predetermined. Predetermination is a consequence of facts about the present and the past. Making  $P(d)$  true by predetermination should involve adaptation of these facts about the present and the past as well.

The question is, however, whether we want to exclude in general for the ontic interpretation function the possibility that it returns possibilities where the truth of  $P(d)$  is predetermined. This is in first instance an empirical question. One might argue that examples of conditionals whose antecedent selects possibilities where the antecedent is predetermined at the utterance time speak against such a proposal (see section 6.3.1).

(128) [I don't know when the train leaves. But]

If the train leaves before noon, then we probably won't catch it.

But we can already account for these examples with the epistemic reading of conditionals, based on the interpretation function *ALearn*.<sup>51</sup> More relevant

---

<sup>51</sup>As will be described in more details in section 6.4.4.9, the epistemic reading of conditionals checks the consequent on those antecedent worlds where it is predetermined that the antecedent formula is true. This is an immediate consequence of calculating, in this case, the hypothetical update with the antecedent using the epistemic update function. This update function selects

for our question would be examples of conditionals the antecedent of which is interpreted to be predetermined at some time in the future. Such examples – if they exist – cannot be captured by the epistemic reading. To model them, we really have to define the interpretation function  $w_p$  of the relevant possibility  $p$  at this point in the future where the antecedent becomes predetermined. Indeed, such examples seem to exist. The following stems from Kaufmann (2005, ex. 47).

- (129) [Let's wait for today's decision regarding his travel arrangements]  
Then, If he arrives tomorrow, we'll book his room tonight.

Kaufmann comments: “This sentence can be paraphrased as ‘*If it is settled (later today) that he arrives tomorrow ...*’ It clearly shows that the Certainty Conditional is part of the interpretation of the antecedent even in predictive conditionals.” (Kaufmann 2005, p. 31) Kaufmann’s predictive conditionals are our ontic readings of conditionals. Hence, Kaufmann concludes from this type of example that the formalization of the ontic reading should allow for predetermination of the antecedent. This suggests the following description of the ontic interpretation function. When applied to a formula  $P(d)$  and a possibility  $p$  it fills the future up to the closest time where the truth of the formula becomes determined. In most cases this will be the time  $g_p(d)$ , because at no earlier time does the truth of  $P(d)$  become determined. But for some formulas it might actually be enough to fill the future up to some point between the times  $t_p$  and  $g_p(d)$ , because at this point the truth of  $P(d)$  becomes predetermined. The problem with this informal description of the ontic reading is that it builds on a theory of predetermination. This is a complex subject that we cannot deal with completely within this dissertation. The best we can do here is to adopt a simplified look on predetermination and define the ontic update function based on it.

In many cases predetermination is a product of the facts about the present and the past plus general deterministic laws. These cases of predetermination can be easily formalized. We need to make a distinction between deterministic and non-deterministic laws. To keep things simple, we will assume that all logical/analytical laws are deterministic, while all causal laws are not deterministic.<sup>52</sup> Now, we impose this notion of determinism as an additional restriction on what a possibility is. A possibility has to be defined for every fact that is predetermined according to the concept of predetermination just described. Furthermore,

---

worlds where the formula it applies to is true. Hence, if the formula is about the future, then worlds are selected where this truth is predetermined.

<sup>52</sup>This is a different notion of determinism than Pearl’s (2000) determinism of causal laws. Pearl assumes that every causal law determines the value of the affected variables for any valuation of the variables representing the causes. In section 5.5.4 we questioned this sort of determinism for causal laws. The description of the ontic reading we proposed in the end does not assume Pearl’s determinism. But now we go a step further. We propose that even in case a causal law determines the value of the effect based on a certain actual valuation of the causes, this does not mean that the world actually behaves as the law predicts it to behave. In other words, we propose that causal laws are default rules; they can be contradicted by the facts.

$p$  cannot be defined for any fact about the future that is not predetermined in this sense. This is how we formalize this idea: a state of affairs  $p$  counts as a possibility if in addition to the conditions of definition 6.4.9 it is also true that  $p$  is defined for  $P$  at some future time  $t > t_p$  if and only if the truth value of  $P$  at  $t$  can be derived from facts about the present and the past of  $p$  and analytical/logical laws.

**6.4.14. DEFINITION.** (Possibilities. An extended definition)

Let  $M = \langle S, \langle T, < \rangle, \langle C, U \rangle, I, now \rangle$  be a model for the language  $\mathcal{L}$  and  $g$  an assignment function.  $S_U \subseteq S$  is the set of all states of affairs  $s$  of  $M$  for which it holds that  $\emptyset \neq \bigcap \{u \in U \mid w_s \subseteq u\} \subseteq w_s$ . A *possibility* for  $M$  and  $g$  is an element  $p \in S_U$ , where (i)  $w_p$  restricted to  $\{t' \mid t' \leq t_p\}$  is a complete function, (ii)  $g_p(d_0) = t_p$ , and (iii) the following condition holds:

$$\begin{aligned} \forall P \in \mathcal{P} \forall t > t_p : \\ \langle P, t \rangle \in \text{dom}(p) \Leftrightarrow [\forall w \in U : [\forall P' \in \mathcal{P} \forall t' \leq t_p : w(P, t') \Leftrightarrow w_p(P', t')] \\ \Rightarrow w(P, t) = w_p(P, t)]. \end{aligned}$$

$W_{M,g} \subseteq S$  is the set of all possibilities with respect to  $M$  and  $g$ .

Given our simple theory of predetermination, a first version of a definition of *AIIntervene* becomes quite straightforward. To model the ontic update of a possibility  $p$  with a consistent formula  $P(d)$  we still use the selection of minimal models. This is necessary, because the ontic update can change the world parameter of a possibility  $p$  even in case the formula  $P(d)$   $p$  is updated with is consistent with the information encoded in the possibility. This is the case when  $p$  is undefined for  $P(d)$ . We define an order that compares similarity with respect to  $p$  and describe the ontic update as selecting maximally similar possibilities that make the formula  $P(d)$  true. For convenience, we repeat the definition of the minimality operation from section 5.6.2.1 of the previous chapter.

**6.4.15. DEFINITION.** (The minimality operator)

Let  $D$  be any domain of objects and  $\leq$  an order on  $D$ . The minimality operator  $Min$  is defined as follows:

$$Min(\leq, D) = \{d \in D \mid \neg \exists d' \in D : d' < d\}$$

We start by considering, as formalization of the similarity relation, simple set inclusion between the world parameters of possibilities.

**6.4.16. DEFINITION.** (The ontic update function *AIIntervene*. A preliminary definition)

Let  $M$  be a model,  $g$  an assignment function,  $p$  a possibility for  $M$  and  $g$ ,  $P \in$

$\mathcal{P}$ , and  $d \in \text{VAR}_i$ . We define  $p =_g p'$  if  $\forall d \in \text{VAR}_i - \{d_0\} : g_p(d) = g_{p'}(d)$ . Furthermore, we define

$$\begin{aligned} \llbracket P(d) \rrbracket_{M,p}^+ &= \{p' \in W_{M,g} \mid p' =_g p \ \& \ w_{p'}(P, g_p(d)) = 1\}, \\ \llbracket P(d) \rrbracket_{M,p}^- &= \{p' \in W_{M,g} \mid p' =_g p \ \& \ w_{p'}(P, g_p(d)) = 0\}. \end{aligned}$$

The ontic update of the possibility  $p$  with the formula  $P(d)$  is defined as follows:

$$\begin{aligned} AIntervene_{M,g}^+(p, P(d)) &= \text{Min}(\subseteq, \{p' \in \llbracket P(d) \rrbracket_{M,p}^+ \mid w_p \subseteq w_{p'}\}), \\ AIntervene_{M,g}^-(p, P(d)) &= \text{Min}(\subseteq, \{p' \in \llbracket P(d) \rrbracket_{M,p}^- \mid w_p \subseteq w_{p'}\}). \end{aligned}$$

According to this definition, *AIntervene* selects minimal possibilities extending  $p$  and making  $P(d)$  true. Minimality is defined simply as set-inclusion between the interpretation functions associated with possibilities. Assume that  $d$  is about the past or present, i.e.  $g_p(d) \leq t_p$ . In this case  $w_p$  is obligatorily defined for  $P(d)$ . If  $w_p(P, g_p(d)) = 1$ , then  $AIntervene_{M,g}^+(p, P(d)) = \{p\}$ . If  $w_p(P, g_p(d)) = 0$ , then  $AIntervene_{M,g}^+(p, P(d)) = \emptyset$ . Similar predictions are made if  $p$  is predetermined for  $P(d)$ . Assume now that  $g_p(d) > t_p$  (i.e.  $d$  refers to the future) and  $p$  is not defined for  $P$  and  $g_p(d)$ . In this case  $AIntervene^+$  selects those possibilities that are defined for the future of  $t_p$  to a time at which the truth of  $P(d)$  becomes determined. That means in particular that we predict that in this case the temporal perspective of the possibility is shifted to this future time. The best way to see how this works is by calculating an example. Let  $\mathcal{P}$  be the set  $\{A, B, C\}$ . As time structure  $T$  we take  $\mathbb{Z}$ . Furthermore, we take the law structure  $\langle C, U \rangle$  where  $U$  is the set of all complete interpretation functions for  $\mathcal{P}$  and  $T$ , and  $C = \langle B, E, F \rangle$  contains two causal laws:  $B = \{A\}$ ,  $E = \{B, C\}$ ,  $F(B) = \langle Z_B, f_B \rangle$  with  $Z_B = \langle A \rangle$  and  $f_B = \{\langle 1, 1 \rangle, \langle 0, 0 \rangle\}$ , and  $F(C) = \langle Z_C, f_C \rangle$  with  $Z_C = \langle B \rangle$  and  $f_C = \{\langle 1, 1 \rangle, \langle 0, 0 \rangle\}$ . Thus, a first causal law says that  $B$  if and only if  $A$ , and a second causal law demands that  $C$  if and only if  $B$ . Now, take the possibility  $p$  with  $t_p = 0$ ,  $g_p(d) = 2$ , and  $w_p$  the function mapping for all times  $t' \leq 0$ ,  $A$ ,  $B$ , and  $C$  to 0 and undefined for all other combinations of times and properties. We will discuss the effects of the ontic update of the possibility  $p$  with the formula  $B(d)$ . The table in figure 6.10 on page 220 describes the interpretation function of a number of possibilities for which we want to know which one(s) end(s) up in  $AIntervene_{M,g}^+(p, B(d))$ . We assume that the assignment associated with these possibilities is the same as the one associated with  $p$ . The value of the temporal perspective of these possibilities follows from the way the interpretation functions are defined.<sup>53</sup> It is clear that  $p$  cannot be in the set  $AIntervene^+(p, B(d))$ , because  $p$  does not make the formula  $B(d)$  true.  $p_1, p_2$ ,

<sup>53</sup>For instance,  $t_{p_4}$  has to be 2, because this is the point at which full determination of  $p_4$  stops.  $t_{p_4}$  cannot be a later time, otherwise  $p_4$  would violate condition (i) of definition 6.4.14. It can also not lie before 2, because in this case it would violate condition (iii) of definition 6.4.14: the model does not contain any analytical/logical laws that allow for predetermination.

and  $p3$  are out, because they are not possibilities according to definition 6.4.14. In the present model there are no deterministic laws that can introduce predetermination for the future. Hence, possibilities with predetermination cannot exist. Thus, we conclude  $AIntervene_{M,g}^+(p, B(d)) \subseteq \{p4, p5, p6, p7\}$ . Because all four possibilities are members of the set  $\{p' \in \llbracket P(d) \rrbracket_{M,p}^+ \mid w_p \subseteq w_{p'}\}$ , we only have to check which of them are minimal according to the order  $\subseteq$ . The way their world parameters are related by set inclusion is given in the graph on the right of figure 6.10 (an arrow points from possibility  $p'$  to possibility  $p''$  if  $w_{p'} \subseteq w_{p''}$ ). As can be read from this graph,  $AIntervene_{M,g}^+(p, B(2)) = \{p4, p5, p6\}$ .

If we take the time-parameter of the possibilities into account as well, we see that the operation  $AIntervene$  shifts the parameter  $t_p$  to the evaluation time of the formula  $AIntervene$  applies to. This is our explanation for the puzzle of the shifted temporal perspective. For most formulas their truth becomes determined just at their evaluation time. In this case we predict that the ontic update function gives back the set of possibilities where the interpreted formula is true and the temporal perspective is the evaluation time of the formula. But it is possible that the truth of a formula becomes determined already before this time. In this case,  $AIntervene$ , as defined so far, will only shift the temporal perspective to this future time before the evaluation time of the formula. This accounts for examples like (129), discussed on page 216.

Unfortunately, the definition of  $AIntervene$  provided in definition 6.4.16 is not fully satisfying. In some cases it does not give an adequate description of the ontic interpretation function. In the following we will develop a more adequate description. But the necessary changes in the definition will raise its complexity considerably. Because the problems with definition 6.4.16 are rather periphrastic for the main enterprise of this chapter, which is to account for the contribution of the tenses to the meaning of conditionals, the reader might want to skip the following excursion on the first reading, jump to section 6.4.4.3 and come back to this point of the text before he reads section 6.4.4.9 on the revision-sensitive version of  $AIntervene$ .

The problems of the definition of the function  $AIntervene$  just mentioned concern the way this function fills up the future to the point where the truth value of the atomic formula the possibility is updated with becomes predetermined. It seems that the given definition allows too many possible futures. First, one might criticize the fact that for the example discussed above  $AIntervene_{M,g}^+(p, B(d))$  contains the possibility  $p6$ . In this possibility a miracle happens at 1 and makes  $B$  true at this time (see figure 6.10). This is totally unrelated to making the antecedent true. Allowing additional unrelated miracles to occur seems to be prohibited for the ontic reading of conditionals. Assume that you believe that a rainy spring causes a large crop in the summer. At the moment it is spring and it is raining a lot. So you expect that we will have a large crop this summer.

	...	$-n$	...	$-2$	$-1$	$0$	$1$	$2$	$3$	...	
$A$	...	0	...	0	0	0	*	*	*	...	$p$
$B$	...	0	...	0	0	0	*	*	*	...	
$C$	...	0	...	0	0	0	*	*	*	...	
$A$	...	0	...	0	0	0	*	*	*	...	$p1$
$B$	...	0	...	0	0	0	*	1	*	...	
$C$	...	0	...	0	0	0	*	*	*	...	
$A$	...	0	...	0	0	0	*	*	*	...	$p2$
$B$	...	0	...	0	0	0	*	1	*	...	
$C$	...	0	...	0	0	0	*	*	1	...	
$A$	...	0	...	0	0	0	*	*	*	...	$p3$
$B$	...	0	...	0	0	0	0	1	*	...	
$C$	...	0	...	0	0	0	*	*	*	...	
$A$	...	0	...	0	0	0	1	0	*	...	$p4$
$B$	...	0	...	0	0	0	0	1	*	...	
$C$	...	0	...	0	0	0	0	0	*	...	
$A$	...	0	...	0	0	0	0	0	*	...	$p5$
$B$	...	0	...	0	0	0	0	1	*	...	
$C$	...	0	...	0	0	0	0	0	*	...	
$A$	...	0	...	0	0	0	1	0	*	...	$p6$
$B$	...	0	...	0	0	0	1	1	*	...	
$C$	...	0	...	0	0	0	0	1	*	...	
$A$	...	0	...	0	0	0	1	0	0	...	$p7$
$B$	...	0	...	0	0	0	0	1	0	...	
$C$	...	0	...	0	0	0	0	0	1	...	

Figure 6.10: An example

Now you consider some conditional with an antecedent not related by laws to the large crop and evaluated at some time after spring – let’s say, what if we go to Spain this year? The theory formulated above would predict that in this case the following conditional is not true.

(130) If we go to Spain this year, we will have a large crop (this summer).

Because the definition of *AIntervene* provided above allows the introduction of new miracles, even though the spring is rainy, the interpretation function can introduce a miracle that will prevent a large crop. We conclude from this example the following: when *AIntervene* fills up the future, then it should do this as far as possible according to the prediction of the laws (even if the laws are not deterministic). No unnecessary miracles should be introduced.

A second problem becomes visible in the fact that the result of the update with  $B(d)$  contains possibility  $p5$ . In  $p5$  the formula  $B(2)$  is made true by a

miracle: even though  $A(1) = 0$ ,  $B$  is set to 1 at time 2. To paraphrase this result, *AItervene* – as defined so far – does not allow for backtracking in the future. But while backtracking is strongly dispreferred if not semantically anomalous for ontic conditionals about the present or the past (see the discussion in chapter 5), it has been argued that nothing similar is true for the future. Compare for instance, (131a) with (131b)

(131) a. If he comes out smiling the interview went well.

b. ??If he had come out smiling, the interview would have gone well.

In chapter 5 we have argued that (131b) becomes acceptable in a context where some convention is present that connects the smiling and the success of the interview. But the relation can also be interpreted as a causal one. In this case, speakers judge (131b) to be not acceptable. However, under the same reading (131a) is fine. Thus, ontic conditionals allow for causal backtracking in case they are about the future.<sup>54</sup> *AItervene* has to be adapted in such a way as to allow for this kind of reasoning. Again, the problem seems to be that so far this function allows for unnecessary miracles. The possibility  $p5$  should be out, because it introduces a miracle to make the formula  $B(d)$  true, although the same effect can also be achieved without a miracle, as possibility  $p4$  illustrates.

To deal with both problems we have to be able to compare how many miracles occur in a possibility. For this purpose we reintroduce the notion of a basis used in Chapter 5 to model the ontic interpretation of conditionals. As done there, the basis of a possibility  $p$  will be defined as the minimal subset of  $w_p$  from which all other facts of  $w_p$  can be derived by law. As proposed in section 5.6.3.1 of the previous chapter we will first introduce the *law closure* of a (partial) interpretation function  $w$ . The law closure describes what can be derived from  $w$  by laws, but only allows for an application of causal laws that reasons from the causes to the effect. We cannot copy definition 5.6.13 exactly from section 5.6.3.1, because we have to make some allowance for the extension of the model with time. But the general outline of the definitions is identical.

**6.4.17. DEFINITION.** (Law closure with time)

Let  $M = \langle S, \langle T, < \rangle, \langle C, U \rangle, I, now \rangle$  be a model with a causal structure  $C = \langle B, E, F \rangle$  where  $B \subseteq \mathcal{P}$  is the set of background variables,  $E = \mathcal{P} - B$  the set of endogenous variables, and  $F$  the function describing the causal dependencies between these variables. We extend interpretation functions  $w$  to intervals  $i \in I(T)$  as follows:

$$w(P, i) = \begin{cases} 1 & \text{if } \forall t \in i : w(P, t) = 1, \\ 0 & \text{if } \forall t \in i : w(P, t) = 0, \\ \text{undefined} & \text{otherwise.} \end{cases}$$

<sup>54</sup>More precisely, the complete process of backtracking has to take place in the future, not only the evaluation of antecedent and consequent.



The *law closure*  $\bar{w}$  of an interpretation function  $w \in S_U$  is the minimal interpretation function  $w' \in S_U$  fulfilling the following conditions.<sup>55</sup>

- (i)  $w \subseteq w'$ ,
- (ii)  $w' = \bigcap \{u \in U \mid w' \subseteq u\}$ , and
- (iii) for all  $P \in E$  with  $Z_P = \langle P_1, \dots, P_n \rangle$  and all  $i \in I(T)$  such that  $w(P, i)$  is undefined the following holds: if there exists  $j \in I(T)$  such that  $j < i$  and  $\neg \exists t \in T(j < t < i)$  and  $f_P(w'(P_1, j), \dots, w'(P_n, j))$  is uniquely determined<sup>56</sup>, then  $w'(P, i)$  is defined and  $f_P(w'(P_1, j), \dots, w'(P_n, j)) = w'(P, i)$ .

The first condition of definition 6.4.17 demands that all facts of  $w$  are also facts of its closure. Condition (ii) requires that the law closure closes under logical/analytical laws. Condition (iii) demands that that endogenous variables of the causal structure whose truth at some interval  $i$  is not already defined in  $w$  obtain in  $\bar{w}$  a truth value at  $i$ , if this truth value is causally determined by other facts of  $\bar{w}$  holding in an interval  $j$  directly preceding  $i$ . Finally, the minimality condition ensures that the closure does not introduce facts not derivable by laws from the facts in  $w$ .

Based on the concept of law closure, we will now introduce the notion of a *basis* of a possibility. This definition is, in principle, identical to definition 5.6.15 of section 5.6.3.1.

**6.4.18. DEFINITION.** (Basis with time)

Let  $M = \langle S, \langle T, < \rangle, \langle C, U \rangle, I, now \rangle$  be a model and  $g$  an assignment function. The *basis*  $b_p$  of a possibility  $p \in W_{M,g}$  is the union of all interpretation functions  $b$  that fulfill the following two conditions: (i)  $b \subseteq w_p \subseteq \bar{b}$  and (ii)  $\neg \exists b' : b' \subseteq w_p \subseteq \bar{b}'$  &  $b' \subset b$ .

With this notion at hand we can now give a definition for an order comparing the similarity of possibilities relative to some possibility  $p$  that overcomes the problems of the simple order of set-inclusion used in definition 6.4.16. Actually, we use two orders. The first order  $\preceq_1^p$  compares similarity with respect to the basis of  $p$ . More precisely, it compares how many miracles have been additionally introduced to those already present in  $p$ . The second order compares similarity additionally with respect to derivable facts. It compares how many of such facts became additionally defined.<sup>57</sup>

<sup>55</sup>The relevant order with respect to which the minimum is calculated is set-inclusion between interpretation functions.

<sup>56</sup>We explicitly want to include here cases where  $w'$  is at  $j$  not defined for all  $P_k$  with  $k \in \{1, \dots, n\}$ , but where those  $k$  for which it is defined are already sufficient to determine from  $f_P$  the value for  $P$  and  $i$ .

<sup>57</sup>There is a close relation between this definition of the order and what has been proposed in Lewis (1979). The definition given here has, among other things, the advantage to not rely on such vague notions as the comparative size of miracles.

**6.4.19. DEFINITION.** (The orders for *AIntervene*)

Let  $M$  be a model,  $g$  an assignment function, and  $p, p_1, p_2 \in W_{M,g}$ . We define for arbitrary interpretation functions  $w$  the function  $w - B$  as the restriction of  $w$  to the domain  $\mathcal{P} - B = E$ , i.e. to property-letters treated as endogenous variables by the causal structure of  $M$ .

$$\begin{aligned} p_1 \preceq_1^p p_2 & \text{ iff } (b_{p_1} - B) - b_p \subseteq (b_{p_2} - B) - b_p, \\ p_1 \preceq_2^p p_2 & \text{ iff } (w_{p_1} - b_{p_1}) - (w_p - b_p) \subseteq (w_{p_2} - b_{p_2}) - (w_p - b_p). \end{aligned}$$

If one compares these orders with those used in the definition of *Intervene* in section 5.6.3.1 of the previous chapter, one finds many similarities, but also some differences. Again, we use two orders, one comparing similarity with respect to the basis, the other with respect to derivable facts. But the way these orders are defined differs. Some of these differences are simplifications we can make because *AIntervene* deals with consistent updates. Some are changes that are necessary to cope with the undefined futures of possibilities.<sup>58</sup> The order comparing similarity of the bases is weaker than the one used in Chapter 5. First, because it does not consider the overlap with the basis of  $p$ , but also because it considers differences between the bases only with respect to newly introduced miracles. The second change is necessary to deal with undefined futures. We need this condition to exclude the problematic predictions concerning backtracking for antecedents about the future that we have discussed above. The second order  $\preceq_2^p$  differs on the first view strongly from  $\leq_2^w$  of section 5.6.3.1. But appearances are misleading. We can neglect comparing the overlap with  $w_p - b_p$  here, because the restriction to consistent updates already ensures that the overlap is total. To compare the extension of the derivable facts that are defined in an alternative possible world did not make so much sense in the framework of Chapter 5, because possible worlds were completely defined interpretation functions. Now, things have changed. Possibilities can be undefined for the future. The second order ensures that the future of an alternative possibility is not defined unnecessarily far.

Analogously to definition 5.6.17 in Chapter 5 we define *AIntervene* by selecting first minima with respect to the first order  $\preceq_1^p$  and then with respect to the second order  $\preceq_2^p$ .

**6.4.20. DEFINITION.** (The ontic update function *AIntervene*. The final definition)

Let  $M$  be a model,  $g$  an assignment function,  $p \in W_{M,g}$ ,  $P \in \mathcal{P}$ , and  $d \in VAR_i$ . The ontic update of the possibility  $p$  with the formula  $P(d)$  is defined as follows:

$$\begin{aligned} AIntervene_{M,g}^+(p, P(d)) &= Min(\preceq_2^p, Min(\preceq_1^p, \{p' \in \llbracket P(d) \rrbracket_{M,p}^+ \mid w_p \subseteq w_{p'}\})), \\ AIntervene_{M,g}^-(p, P(d)) &= Min(\preceq_2^p, Min(\preceq_1^p, \{p' \in \llbracket P(d) \rrbracket_{M,p}^- \mid w_p \subseteq w_{p'}\})). \end{aligned}$$

<sup>58</sup>In Chapter 5 possible worlds where completely defined interpretation functions and the difference between past and future was non-existent.

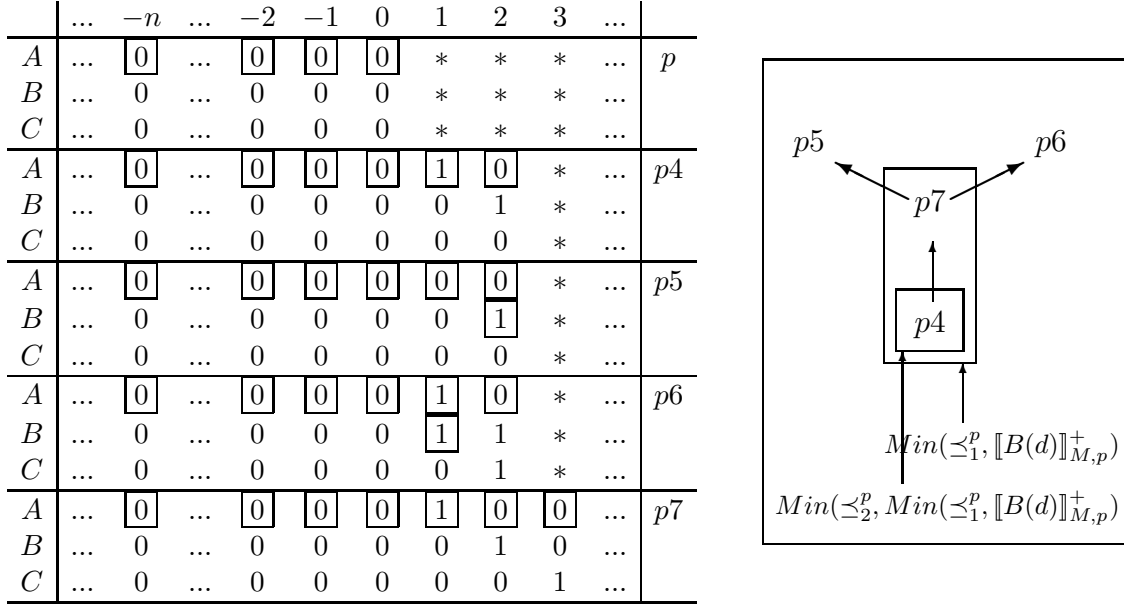


Figure 6.11: The example again with the new orders

It is easy to see that the predictions made by this definition of *AIntervene* are the same as before when it comes to formulas whose truth value is already defined for  $p$ . If this is not the case, then the predictions differ. For illustration, let us apply *AIntervene* to the formula  $B(d)$  with the respect to the language and the model introduced when we discussed the working of the preliminary version of *AIntervene*. We consider again the possibilities described in figure 6.10, page 220. As before,  $p$  is not in  $AIntervene_{M,g}^+(p, B(d))$ , because it does not make the formula true, and  $p1$ ,  $p2$ , and  $p3$  are out, because they are not possibilities according to definition 6.4.14. But how do the new orders  $\preceq_1^p$  and  $\preceq_2^p$  relate the remaining four possibilities? To calculate this, we first need to know the bases of each of the four possibilities. In figure 6.11 the elements of the bases are marked by boxes around the relevant entries in the truth table. For convenience, also the basis of  $p$  is given. In the graph on the right the way the orders  $\preceq_1^p$  and  $\preceq_2^p$  relate the possibilities  $p4$ ,  $p5$ ,  $p6$ , and  $p7$  is described. A thick arrow points from possibility  $p_i$  to possibility  $p_j$  if  $p_i \prec_1^p p_j$ , a thin arrow represents the order relation  $\preceq_2^p$ . The second order is only marked for the elements minimal with respect to the first order. As the graph shows, the possibilities  $p5$  and  $p6$  are suboptimal according to the first order. The reason is that they introduce unnecessarily many miracles. The possibility  $p7$  is not minimal according to the second order, because it is defined unnecessarily far into the future.  $p4$  comes out as the unique minimum – as intended –, i.e.  $AIntervene_{M,g}^+(p, B(d)) = \{p4\}$ .

Let us further illustrate the way *AIntervene* works with a more general state-

ment. Assume that the atomic formula  $P(d)$  *AIIntervene* applies to is not expected to be false in the possibility  $p$  it takes as argument. That means  $P(d)$  can be made true in  $p$  without additional law-violations. We will show that, given some additional restrictions on the law structure of the model, in this case the possibilities in  $AIIntervene_{M,g}^+(p, P(d))$  follow the expectations of  $p$ . We start by defining what we mean by saying that a possibility  $p'$  *follows the expectations* of  $p$ . This is the case if  $w_p'$  extends  $w_p$ , but does not introduce new miracles on the way. As a consequence, we have  $\overline{w_p} \subseteq \overline{w_{p'}}$ .

**6.4.21. DEFINITION.** (Following the expectations)

Let  $M$  be a model,  $g$  an assignment function,  $p, p' \in W_{M,g}$ . We say that  $p'$  *follows the expectations of  $p$*  ( $p \rightsquigarrow p'$ ), if the following three conditions are fulfilled: (i)  $w_p \subseteq w_{p'}$ , (ii)  $p' =_g p$ , and (iii)  $(b_{p'} - B) = b_p - B$ .

Such possibilities  $p'$  that follow the expectations of  $p$  do not have to exist. Whether they do depends on the way the law system is set up. However, if they exist, then these are the possibilities chosen by *AIIntervene*.<sup>59</sup>

**6.4.22. FACT.** Let  $M$  be a model where  $U$  is the set of all complete interpretation functions for  $\mathcal{P}$ . Let  $g$  be an assignment function,  $p \in W_{M,g}$ ,  $P \in \mathcal{P}$  and  $d \in VAR_i$ . Assume that  $\exists p' \in \llbracket P(d) \rrbracket_{M,p}^+ : p \rightsquigarrow p'$ . Then the following equation holds.

$$AIIntervene_{M,g}^+(p, P(d)) = \text{Min}(\preceq_2^p, \{p' \in \llbracket P(d) \rrbracket_{M,p}^+ \mid p \rightsquigarrow p'\}).$$

In this formulation the result is restricted to models the law structure of which does not know any analytical laws.

Before we can conclude this section, we first have to say what *AIIntervene* does on the level of cognitive states. We define the ontic update of a cognitive state  $c$  with an atomic formula  $P(d)$  as the union of the results we obtain by applying *AIIntervene* to  $P(d)$  and every possibility in  $c_{\eta(c)}$ .

**6.4.23. DEFINITION.** (Intervention for cognitive states)

Let  $M$  be a model,  $g$  an assignment function,  $c$  be a cognitive state,  $P \in \mathcal{P}$ , and  $d \in VAR_i$ . The *AIIntervene*-update of  $c$  with  $P(d)$  is defined as follows:

$$\begin{aligned} AIIntervene_{M,g}^+(c, P(d)) &= c[\eta(c) / \bigcup_{p \in c_{\eta(c)}} AIIntervene_{M,g}^+(p, P(d))], \\ AIIntervene_{M,g}^-(c, P(d)) &= c[\eta(c) / \bigcup_{p \in c_{\eta(c)}} AIIntervene_{M,g}^-(p, P(d))]. \end{aligned}$$

<sup>59</sup>The proof is left to the reader. Notice, that the fact does not hold if we replaced the condition  $\exists p' \in \llbracket P(d) \rrbracket_{M,p}^+ : p \rightsquigarrow p'$  by the condition  $\overline{w_p}(P, g_p(d)) \neq 0$ . In this case there can be possibilities in  $AIIntervene_{M,g}^+(p, P(d))$  that do not follow the expectations of  $p$ . Things change if we strengthen the order  $\preceq_2^p$  a bit. We formalized the derivable facts of a possibility  $p$  as the difference  $w_p - b_p$ . One might suggest that instead of only counting those facts derivable from the basis that  $w_p$  is defined for, we also take the derivable facts into account that  $w_p$  is not defined for. Hence, we use  $\overline{w_p} - b_p$ . Which of the two formalizations should be chosen is in the end an empirical question, but one that we do not want to decide without a more serious study of the data. However, strengthening the order  $\preceq_2^p$  in the proposed way would allow us to prove the fact for the weaker condition  $\overline{w_p}(P, g_p(d)) \neq 0$ .

### 6.4.4.3 Support and enforcement

Before we come to the meaning of the other items of the lexicon of  $\mathcal{L}$ , let us first introduce two important notions for the further discussion.

*ALearn* describes the update with a formula  $\psi$  that is taken to convey information about the actual world. Sometimes, the information that  $\psi$  conveys may already be present in a basic state. For atomic sentences this is, for instance, the case if  $\psi$  is true in every possibility in this basic state. Because the epistemic update works by eliminating possibilities, we can test whether the information  $\psi$  conveys is already contained in the basic state  $c_{\eta(c)}$  of some cognitive state  $c$  by checking whether all possibilities are still there after updating  $c$  with  $\psi$ . But we have to be careful. It may be the case that the formula conveys dynamic information and eliminates a possibility because of properties of its assignment function. Therefore, we rather demand that for all possibilities  $p$  in  $c_{\eta(c)}$  we can find another possibility  $p'$  after update that only differs from  $p$  with respect to the dynamic information it encodes. Thus, we demand that the world parameter is the same ( $w_p = w_{p'}$ ), the temporal perspective is the same ( $t_p = t_{p'}$ ), but the assignments may differ in all variables except  $d_0$  ( $g_p(d_0) = g_{p'}(d_0)$ ).

#### 6.4.24. DEFINITION. (Support)

Let  $M$  be a model,  $g$  an assignment function,  $c$  a cognitive state, and  $\psi$  a formula of  $\mathcal{L}$ . A possibility  $p \in W_{M,g}$  *subsists* in a possibility  $p' \in W_{M,g}$  ( $p \rightarrow p'$ ), if  $w_p = w_{p'}$  &  $t_p = t_{p'}$  &  $g_p(d_0) = g_{p'}(d_0)$ . We define that  $c$  *supports*  $\psi$  ( $c \models \psi$ ), if  $\forall p \in c_{\eta(c)} \exists p' \in (ALearn_{M,g}^+(c, \psi))_{\eta(c)} : p \rightarrow p'$ .

A similar notion can be defined for the ontic update function *AIntervene*. However, this update function is not about learning new information  $\psi$ , but about making  $\psi$  true. The *AIntervene*-analogue to that  $c_{\eta(c)}$  already contains the information conveyed by  $\psi$  is that  $c_{\eta(c)}$  already expects this change to  $\psi$  to be happening. Or, in other words, if the causal laws are obeyed, then every possibility in  $c_{\eta(c)}$  will naturally develop into one where  $\psi$  is true. In this case we say that a cognitive state  $c$  *forces* the formula  $\psi$ .

A first idea of how to formalize *enforcement* is to work in analogy with definition 6.4.24 and insert the relation *following the expectations* where we used subsistence before.

#### 6.4.25. DEFINITION. (Enforcement. A preliminary definition)

Let  $M$  be a model,  $g$  an assignment function,  $c$  a cognitive state, and  $\psi$  a formula of  $\mathcal{L}$ . We define that a cognitive state  $c$  *forces*  $\psi$  ( $c \models \psi$ ), if  $\forall p \in c_{\eta(c)} \exists p' \in (AIntervene_{M,g}^+(C, \psi))_{\eta(c)} : p \rightsquigarrow p'$ .

But what we would express this way is rather that  $\psi$  is consistent with the expectations of  $c$ , not that it is actually entailed by them. Therefore, we add a second condition: all possibilities  $p'$  in the update with  $\psi$  that follow  $p$  make  $\psi$

true in this part of their interpretation function that agrees with the expectations of  $p$ .

**6.4.26. DEFINITION.** (Enforcement. The final definition)

Let  $M$  be a model,  $g$  an assignment function,  $c$  a cognitive state, and  $\psi$  a formula of  $\mathcal{L}$ . We define that a cognitive state  $c$  *forces*  $\psi$  ( $c \models \psi$ ), if

- (i)  $\exists p' \in (AIntervene_{M,g}^+(c, \psi))_{\eta(c)} : p \rightsquigarrow p'$ , and
- (ii)  $(\bigcap \{w_{p'} \mid p' \in (AIntervene_{M,g}^+(c, \psi))_{\eta(c)}\}) \subseteq \overline{w_p}$ .

**6.4.4.4 The meaning of the basic logical operators**

Next, we provide the update rules for the operators  $\wedge$ ,  $\vee$ , and  $\neg$ . As explained earlier, negation is in this framework interpreted as changing the polarity of the update function: the positive update with a formula  $\neg\psi$  results in the negative update with  $\psi$ , and vice versa. The standard definition, letting the update with a negative sentence result in the basic state that contains all possibilities lost in the update with the unnegated sentence, makes wrong predictions in case the meaning of formulas in scope of the negation can be undefined. The meanings assigned to conjunction and disjunction conform to common practices in dynamic semantics.<sup>60</sup>

**6.4.27. DEFINITION.** (The basic logical operators)

Let  $M$  be a model,  $c$  be a cognitive state, and  $\psi, \phi$  formulas of  $\mathcal{L}$ .

- (i)  $c[\neg\psi]_{M,g}^+ = c[\psi]_{M,g}^-$ ,  
 $c[\neg\psi]_{M,g}^- = c[\psi]_{M,g}^+$ .
- (ii)  $c[\psi \wedge \phi]_{M,g}^+ = c[\psi]_{M,g}^+[\phi]_{M,g}^+$ ,  
 $c[\psi \wedge \phi]_{M,g}^- = c[\psi]_{M,g}^- \cup c[\phi]_{M,g}^-$ .<sup>61</sup>
- (iii)  $c[\psi \vee \phi]_{M,g}^+ = c[\psi]_{M,g}^+ \cup c[\phi]_{M,g}^+$ ,  
 $c[\psi \vee \phi]_{M,g}^- = c[\psi]_{M,g}^-[\phi]_{M,g}^-$ .

The way we treat conjunctions is not entirely satisfying. As said earlier, the approach developed here makes very specific predictions for intra-sentential modal subordination: the reference context is the last element the output cognitive state of the last update is defined for. In the case of conjunction this means that if the

---

<sup>60</sup>There are some open questions concerning the predictions made by this approach for anaphorical relations. In particular, one may doubt the correctness of using set union for the negative update with a conjunction and the positive update with a disjunction. We will not discuss these issues here. They are only of marginal relevance for the central subject of our research.

<sup>61</sup>The union of two cognitive states is defined as the cognitive state resulting from taking the union of the basic states with the same index.

first conjunct introduces a new subordinate context, this context is the basic state the second conjunct is updated to. This is in general not correct. Intuitively, the reference context for the second conjunct is normally the same as for the first conjunct. We will come back to this point in the discussion, section 6.5.

#### 6.4.4.5 The meaning of the temporal operators.

We distinguish two tenses for English: a present tense and a past tense. For their interpretation we adopt a referential analysis, according to which tenses do not quantify over times, but express anaphorical relations to earlier introduced times in the context (see, for instance, Partee 1973, Enč 1986, Kamp & Reyle 1993 and Kratzer 1998). We follow Heim (1994) in assuming that tenses come with presuppositions about what a proper anaphoric referent is: the meaning of a tensed formula is only defined if there is a time available in the context that fulfills the conditions the tense imposes. Heim & Kratzer (1998) propose that in this respect the tenses behave similar to gender features of pronouns. However, we go even further in proposing that the tense *is* a temporal pronoun.<sup>62</sup> They have the analogue type and meaning standardly assumed for individual pronomina like *she* and *he*. Hence, the tenses are treated as variables whose interpretation is determined by the assignment function.

A distinguishing aspect of temporal pronouns is that the restrictions on the location of the referent they come with are relative to another time. This is the temporal perspective  $t_p$  of the possibilities  $p$  in the basic state a tensed formula is updated to. A past tense pronoun presupposes that the assignment function maps it to some time before  $t_p$ , a present tense pronoun presupposes that the assignment function maps it to some time identical to  $t_p$  or in the future of  $t_p$ . In most cases the temporal perspective is the utterance time. In subordinated contexts, however, what counts as the *now*, the present time, may be shifted to the future of the utterance time. In relating the interpretation of the tenses to the temporal perspective instead of the utterance time, we claim that tenses are not in a strict sense deictic.

#### 6.4.28. DEFINITION. (The interpretation rules of the tenses)

Let  $M$  be a model,  $g$  an assignment function,  $c$  a cognitive state, and  $\psi$  an expression of type  $[i]$ . The update of  $c$  with  $PRES_n(\psi)$  is defined only if  $\forall p \in c_{\eta(c)} : g_p(d_n) \geq t_p$ . Where defined,

$$\begin{aligned} c[PRES_n(\psi)]_{M,g}^+ &= c[\psi(d_n)]_{M,g}^+, \\ c[PRES_n(\psi)]_{M,g}^- &= c[\psi(d_n)]_{M,g}^-. \end{aligned}$$

The update of  $c$  with  $PAST_n(\psi)$  is defined only if  $\forall p \in c_{\eta(c)} : g_p(d_n) < t_p$ . Where

---

<sup>62</sup>This step may become superfluous after event semantics is introduced.

defined,

$$\begin{aligned} c[PAST_n(\psi)]_{M,g}^+ &= c[\psi(d_n)]_{M,g}^+, \\ c[PAST_n(\psi)]_{M,g}^- &= c[\psi(d_n)]_{M,g}^-. \end{aligned}$$

The meaning proposed here for the present tense follows approaches like Kaufmann (2005) and others in letting the present tense refer to the present as well as to the future. Even though in principle we concede to the present tense to refer to a future time, we can account for the observation that simple sentences marked with the present tense rarely refer to the future and if they do, the truth of the sentence is claimed to be predetermined at the utterance time. Remember that possibilities are associated with only partially defined interpretation functions. In particular, possibilities are only defined for the future, if this aspect of the future is predetermined by deterministic laws. Only few facts are predetermined in this sense. Thus, statements of the form  $PRES_n(P)$  where  $d_n$  is mapped to some time in the future normally do not result in a successful epistemic update, because the value of  $P(d_n)$  is undefined in the possibilities of the basic state the update is applied to. If the update is successful, then this is so because the truth of  $P(d_n)$  is predetermined at the utterance time.

#### 6.4.4.6 The meaning of the perfect

The meaning of the perfect proposed here strongly simplifies matters. This simplification is a necessary consequence of our decision not to get involved in event semantics at the present stage of work and, therefore, ignore all matters of aspect.<sup>63</sup>

We will assign a relational temporal meaning to the perfect. We propose that the perfect introduces a new discourse referent for times, maps it to some time before its evaluation time and demands that the condition in its scope is true at that time. The definition of the negative update with the perfect claims that for all times before the evaluation time  $d$  of the perfect the truth value of the formula  $\psi$  in the scope of the perfect is defined and false.<sup>64</sup>

The form of the interpretation rule of the perfect given below would have looked much less complex, if we had introduced existential quantification over times and the order relation between times into the formal language and had given separate interpretation rules for them. But because there is no motivation in the form of English (conditional) sentences for distinguishing these expressions, we abstained from doing so.

---

<sup>63</sup>To be precise, we should have added above that we also ignore aspectual properties of the simple present and the simple past.

<sup>64</sup>An alternative definition would have been to demand that for all times before the evaluation time  $\psi$  is not true. However, the first formulation appears to be more in accordance with intuitions.



**6.4.29. DEFINITION.** (The interpretation rule of the perfect)

Let  $M$  be a model,  $g$  an assignment function,  $c$  a cognitive state,  $\psi$  an expression of type  $[i]$ , and  $d \in VAR_i$ . For  $d_1, d_2 \in VAR_i$  we define  $c[d_1 < d_2] = c[\eta(c)/\{p \in c_{\eta(c)} \mid g_p(d_1) < g_p(d_2)\}]$ . For  $t \in T$  we define  $c[d_1/t] = c[\eta(c)/\{p[d_1/t] \mid p \in c_{\eta(c)}\}]$ .

$$\begin{aligned} c[PERF_n(\psi)(d)]_{M,g}^+ &= \bigcup_{t \in T} c[d_n/t][d_n < d][\psi(d_n)]_{M,g}^+, \\ c[PERF_n(\psi)(d)]_{M,g}^- &= c[\eta(c)/\{p \in c_{\eta(c)} \mid \\ &\quad \forall t \in T \exists p' \in c[d_n/t][d_n < d][\psi(d_n)]_{M,g}^-_{\eta(c)} : p \rightarrow p'\}]. \end{aligned}$$

**6.4.4.7 The meaning of the modals**

The treatment of the modal operators *WOLL* and *MOLL* proposed here focusses on the uses of the corresponding modals in conditionals. Furthermore, we will only account for non-root meanings in the context of conditionals. Eventually, the approach should be extended to their meaning in other contexts and other modals in conditionals as well.

The basic intuition about the meanings of *WOLL* and *MOLL* that we will try to capture here is the following. A sentence *WOLL*  $\psi$  is accepted by a cognitive state  $c$ , if  $\psi$  is expected in  $c_{\eta(c)}$ ; a sentence *MOLL*  $\psi$  is accepted by  $c$ , if  $\psi$  is possible in  $c_{\eta(c)}$ .<sup>65</sup> Thus, following approaches like Veltman (1996) these modal claims are interpreted as performing tests on the context that is updated with them. There is a straightforward way to formalize this idea for the case of *MOLL*. In order to see whether  $\psi$  is possible with respect to a basic state  $c_{\eta(c)}$  of a cognitive state  $c$ , we test whether the intervention in  $c$  with  $\psi$  does not lead to the absurd cognitive state. This results in the following interpretation rule for *MOLL* ( $c$  is a cognitive state,  $\psi$  an expression of type  $[i]$ , and  $d, d_n \in VAR_i$ ).

$$\begin{aligned} c[MOLL_n(\psi)(d)]_{M,g}^+ &= \begin{cases} AIntervene_{M,g}^+(c[\eta(c)], \psi(d_n)) \text{ if } AIntervene_{M,g}^+(c[\eta(c)], \psi(d_n)) \notin \perp \\ c[\eta(c)/\emptyset] \text{ if } AIntervene_{M,g}^+(c[\eta(c)], \psi(d_n)) \in \perp \end{cases} \end{aligned}$$

This rule can be simplified as follows.<sup>66</sup>

$$c[MOLL_n(\psi)(d)]_{M,g}^+ = AIntervene_{M,g}^+(c[\eta(c)], \psi(d_n))$$

Next, we come to the meaning of *WOLL*( $\psi$ ). To capture the intuitive description of the meaning of this formula provided above, we have to formalize the idea that the formula  $\psi$  in the scope of *WOLL* is expected in the basic state  $c_{\eta(c)}$ . Expectations are taken here to be what can be inferred from a possibility by the

<sup>65</sup> *WOLL*  $\psi$  and *MOLL*  $\psi$  are not sentences of our formal language  $\mathcal{L}$ , but at this point we only want to give a general idea of the approach proposed here.

<sup>66</sup> The simplified version does not always return the same result as the original rule. In case  $AIntervene_{M,g}^+(c[\eta(c)], \psi(d_n))$  is empty, the first rule sets  $c_{\eta(c)}$  to the empty set, while the second rule does the same with  $c_{\eta(c[\eta(c)])}$ . In both cases the result is an absurd cognitive state.

laws. There may be much more going into the calculation of real life expectations than this, but the framework is open to extensions on this point.<sup>67</sup> With this concept of expectations the basic intuition about the meaning of *WOLL* can be formalized using the notion of enforcement (see definition 6.4.26) introduced earlier ( $c$  is a cognitive state,  $\psi$  an expression of type  $[i]$ , and  $d, d_n \in VAR_i$ ).

$$c[WOLL(\psi)(d)]_{M,g}^+ = \begin{cases} AIntervene_{M,g}^+(c[\eta(c)], \psi(d_n)) & \text{if } c \models \psi(d_n) \\ c[\eta(c)/\emptyset] & \text{if } c \not\models \psi(d_n) \end{cases}$$

So far, we have only given positive update rules for the modals. For the interpretation of negation their negative counterparts are needed as well. Because the update rules for the modals are formulated as test conditions and these test conditions can only have two outcomes, we can use standard dynamic negation and define the negative update with modal formulas as the (dynamic) complement of the positive update.

**6.4.30. DEFINITION.** (Two-valued dynamic negation)

Let  $M$  be a modal, and  $c$  and  $c'$  cognitive states. We define  $c \div c'$  to be the cognitive state  $c''$  where  $c''$  is like  $c$  except that  $c''_{\eta(c)} = \{p \in c_{\eta(c)} \mid p \not\models c'_{\eta(c)}\}$ .

Finally, we have to say something about the temporal properties of the modals. We propose that besides their modal force, *WOLL* and *MOLL* also have an anaphorical, temporal meaning component. The modals evaluate the phrases in their scope at some earlier introduced time that has to be in the future of the evaluation time of the modal. This treatment differs, for instance, from what has been proposed in Condoravdi (2002). She claims that the modals do not involve a temporal anaphor, but rather an operation opposite to the perfect that existentially quantifies over times in the future of the evaluation time of the modal. This, however, seems not to be in accordance with the intuitions about the temporal properties of the modals. If you say *it might rain*, then there is a concrete interval that you have in mind in which you think that it is possible that it rains. The sentence is not true simply because it will rain at some time. So, the problem with existential quantification over time in modals is the same as that noticed by Partee for an existential treatment of the simple past with her famous stove example. With all this said, we can now properly define the update conditions for the modalities.

**6.4.31. DEFINITION.** (The ontic interpretation rule of the modals)

Let  $M$  be a model,  $g$  an assignment function,  $c$  be a cognitive state,  $\psi$  an expres-

---

<sup>67</sup>Other information that might be used to calculate expectations are, for instance, statistical laws.

sion of type  $[i]$ , and  $d, d_n \in VAR_i$ .

$$\begin{aligned}
c[MOLL_n(\psi)(d)]_{M,g}^+ &= \begin{cases} AIntervene_{M,g}^+(c[\eta(c)], \psi(d_n)) \text{ if } \forall p \in c_{\eta(c)} : g_p(d_n) \geq g_p(d) \\ \text{undefined otherwise,} \end{cases} \\
c[MOLL_n(\psi)(d)]_{M,g}^- &= c \div c[MOLL_n(\psi)(d)]_{M,g}^+, \\
c[WOLL_n(\psi)(d)]_{M,g}^+ &= \begin{cases} AIntervene_{M,g}^+(c[\eta(c)], \psi(d_n)) \text{ if } c \models \psi(d_n) \\ c[\eta(c)/\emptyset] \text{ if } c \not\models \psi(d_n) \\ \text{undefined if } \neg \forall p \in c_{\eta(c)} : g_p(d_n) \geq g_p(d), \end{cases} \\
c[WOLL_n(\psi)(d)]_{M,g}^- &= c \div c[WOLL_n(\psi)(d)]_{M,g}^+.
\end{aligned}$$

An interesting property of this approach is that we do not need to distinguish an epistemic reading and an ontic reading for these modalities. The epistemic reading falls out if all possibilities are defined for the formula in scope of the modal. What we cannot have so far is an epistemic reading for formulas about the future the truth conditions of which are not defined in every possibility. There are some observations in Crouch (1993) that speak for the distinction of an epistemic reading of conditionals. We will come back to them below. They motivate the following definition.

**6.4.32. DEFINITION.** (The epistemic interpretation rule of the modals)

Let  $M$  be a model,  $g$  an assignment function,  $c$  be a cognitive state,  $\psi$  an expression of type  $[i]$ , and  $d, d_n \in VAR_i$ .

$$\begin{aligned}
c[MOLL_2(\psi)(d)]_{M,g}^+ &= \begin{cases} ALearn_{M,g}^+(c[\eta(c)], \psi(d_n)) \text{ if } \forall p \in c_{\eta(c)} : g_p(d_n) \geq g_p(d) \\ \text{undefined otherwise} \end{cases} \\
c[MOLL_2(\psi)(d)]_{M,g}^- &= c \div c[MOLL_2(\psi)(d)]_{M,g}^+, \\
c[WOLL_2(\psi)(d)]_{M,g}^+ &= \begin{cases} ALearn_{M,g}^+(c[\eta(c)], \psi(d_n)) \text{ if } c \models \psi(d_n) \\ c[\eta(c)/\emptyset] \text{ if } c \not\models \psi(d_n) \\ \text{undefined if } \neg \forall p \in c - \eta(c) : g_p(d_n) \geq g_p(d) \end{cases} \\
c[WOLL_2(\psi)(d)]_{M,g}^- &= c \div c[WOLL_2(\psi)(d)]_{M,g}^+.
\end{aligned}$$

To illustrate the working of the semantics for the modals suggested here, let us discuss some of the interesting predictions it makes. To start with, assume that both the evaluation time for the modal and the time at which the formula in the scope of the modal is evaluated, lie in the past. This is a possible interpretation if the modal is marked for the past tense (our lexicon excludes a past interpretation of *might*) or the modal stands in the scope of the perfect (not possible in our syntax, but one might extend the syntax in this respect, see for instance Condoravdi 2002). In this case, we predict that an update with a sentence  $MOLL_n(\psi)(d)$  is successful, if  $\psi(d_n)$  is epistemically possible in the basic state to which the formula is updated.  $WOLL_n(\psi)(d)$  is successfully updated, if  $\psi(d_n)$  is epistemically necessary in this basic state. This prediction is independent of whether the modals are interpreted according to their ontic reading (definition 6.4.31) or their epistemic reading (definition 6.4.32). An important difference with approaches

to the meaning of the modals like Condoravdi (2002) is that also when evaluated in the past, the modals do not quantify over possibly counterfactual alternatives. In Condoravdi (2002), for instance, the ontic interpretation quantifies roughly over all ways how the future may turn out to be. If the modal is evaluated in the past, then the quantification ranges over all possible futures at this point in the past. This set can contain alternative possibilities that are counterfactual. A clear advantage of our treatment is that – in contrast to Condoravdi (2002) – it can account for the intuition that a sentence (132) can be true even if at the time of birth it was not yet predetermined that the child would become a king.

(132) A child was born that would be king.

But at the same time, this approach also predicts that *It might have been  $\psi$* , that may be analyzed as having the logical form  $PRES_l(PERF_m(MOLL_n(\psi)))$ <sup>68</sup> as statement about the actual world cannot be about a counterfactual possibility. In the approach of Condoravdi (2002) this is possible. Under its ontic reading the only thing needed is some point in the past where the world could still have developed in a way such that  $\psi$  becomes true. The approach of Condoravdi seems to win in this point because we can say things like (133a) and (133b).

(133) a. He might have won the game, but in the end he didn't.

b. It might have rained this day. Fortunately, it didn't.

But notice, that the semantics assigned to these sentences by Condoravdi (2002) is not convincing, because it predicts that for nearly any false  $\psi$  an update with *might have  $\psi$*  is successful – there will always be some point in the past where the semantic value of  $\psi$  is not determined yet.<sup>69</sup> Furthermore, we can say that the examples given above are no problem for the approach proposed here, because such sentences in their counterfactual reading are never uttered out of the blue. The speaker always has some condition in mind that would have brought about the described possibility. Hence, such sentences refer always to some hypothetical basic state derived from the information state about the actual world by revision.

A very uncommon prediction of the ontic reading for the modalities proposed here is that if the evaluation time for the formula in the scope of the modals lies in the future, then the temporal perspective of the subordinated basic state introduced by the modal is shifted to this time in the future. This certainly allows one to account for the variant of the puzzle of the shifted temporal perspective for modals. However, one might wonder whether this prediction is also convincing for cases of modal subordination. In the section where we described the model

<sup>68</sup>We ignore mood-markings for the moment and assume that the perfect can scope over modals.

<sup>69</sup>A possible solution may be to motivate contextual restrictions on the evaluation time of a modal, even in case it occurs in scope of the perfect.

(section 6.4.3) we sketched a very preliminary approach to account for modal subordination. As is standardly proposed, modal subordination is explained here as reference to some previously introduced hypothetical state. The proposed semantics for the modals predicts that in case  $\psi$  is updated to some hypothetical basic state introduced by some modal statement about the future, the tenses in  $\psi$  are interpreted with respect to a future-shifted reference time. Indeed, something along these lines is needed to explain the observation that the perfect is obligatory for backtracking reasoning after ontic modal statements about the future.

- (134) a. John might come out smiling.  
       b. \*(In this case) the interview would go well.  
       c. (In this case) the interview would have gone well.

Notice that the proposed epistemic reading of the modals does not predict such a future-shift of the temporal perspective of the updated possibilities. On the other hand, it is also not difficult to find examples for modal subordination, where the reference time for the tenses in the second sentence is not shifted (see the example below).<sup>70</sup>

- (136) a. (I don't know what John has decided.) He might take the train tonight.  
       b. (In this case) He would buy the tickets this afternoon.

This is not a very strong argument in favor of an epistemic reading of conditionals. The context of (136b) implies that John has already made his decision, i.e. it is determined at the utterance time whether John takes the train tonight. But in this case already the proposed ontic reading of the modals would account for the example. Stronger support for the epistemic reading comes from an observation of Crouch (1993) (see section 6.3.1). He observes that the shift of the temporal perspective in the scope of modals is not obligatory. The example he uses to illustrate his observation ((124a) on page 191) is not placed in a context where the phrase in the scope of the modal is known to be predetermined. If he is right with his observation, then we would have an empirical argument in favor of the epistemic reading.

---

<sup>70</sup>As an aside, after the semantics for the moods is introduced in the next section it will become clear that sentence (135b) in its indicative reading cannot refer to the subordinated context introduced by the sentence (135a) below. The subjunctive reading is semantically anomalous for the same reasons as was (134b) as continuation of (134a).

- (135) a. John might be here by tomorrow noon.  
       b. \*(In this case) he took the train leaving 9 pm this evening.

A final distinguishing property of this approach to the modalities that we will discuss here is that the proposed meanings for *MOLL* and *WOLL*, while related, are not duals of each other. Therefore, it is possible to successfully update a cognitive state with a sequence  $MOLL_n(\psi)(d) \wedge WOLL_n(\neg\psi)(d)$ . The reason is that *AIntervene* can change the future contrary to what is predicted by causal laws. Even in case the causal laws predict  $\neg\psi(d_n)$  (and, therefore an update with  $WOLL_n(\neg\psi)(d)$  is successful) the future still might be undecided for  $\psi(d_n)$  (and, therefore, an update with  $AIntervene_{M,g}^+(c, \psi(d_n))$  might be successful). However, this approach predicts that an update with  $MOLL_n(\psi)(d) \wedge WOLL_n(\neg\psi)(d)$  can only go through, if  $\psi(d_n)$  goes against the predictions of the causal laws. Together with the semantics proposed for the moods in the next section, this would mean that the modal *MOLL* in this formula has to be realized as *might* and not as *may*. Actually, this appears to be confirmed by intuitions, as the next example shows.

(137) It may<sup>?</sup>/might happen, but it will not happen.

Sentences like this are notoriously problematic for many approaches to the meaning of these modals, particularly those that treat *WOLL* and *MOLL* as duals of each other. Our approach has no problems with accounting for it.

#### 6.4.4.8 The meaning of the moods

We propose that all (assertive) English sentences are marked for mood. Three moods are distinguished: an indicative mood, a subjunctive mood, and a counterfactual mood. The basic idea for the role the mood plays for interpretation is simple: it tells the interpreter something about how the update of a cognitive state  $c$  with a formula  $\psi$  relates to the beliefs encoded in  $c$  about the actual world, i.e. the information present in  $c[\psi]_0$ . The mood, thereby, helps to select the proper subordinate state to update the formula to. The indicative mood and the subjunctive mood are about the expectations of  $c[\psi]_0$ . Already in the last section, where the semantics of the modals were discussed, we introduced a formalization of expectations within the present framework: the expectations of a possibility  $p$  are facts that can be derived from  $p$  by general laws; the expectations of a basic state are what is expected in everyone of its possibilities.<sup>71</sup> The difference between the way expectations are involved in the semantics of *WOLL* and the meaning of the moods is that *WOLL* depends on the expectations in the basic state  $c_{\eta(c)}$  where the update takes place. This may very well be some subordinate, hypothetical state. The mood, however, always refers to the expectations of the maximally superordinated basic state of a cognitive state,  $c_0$ .

We propose that the indicative mood checks whether the update is consistent with the expectations in  $c_0$  (after update). This is formalized by the condition

---

<sup>71</sup>As said before, this gives only a partial picture of what goes into our expectations.

that after update every possibility in the last basic state the updated cognitive state  $c$  is defined for follows some possibility in its basic state  $c_0$ . That means that every possibility in the (hypothetical) state introduced last has to be identical to, or the future of what might according to  $c$  be, the actual world. If this is the case, then the formula in the scope of the mood is indeed updated to the cognitive state. Otherwise the interpretation process breaks down. It is important to check the expectations in  $c_0$  after the update. It may be the case that the updated formula gives information about the actual world that does not conform to expectations, but is consistent with what is known. For such statements still the indicative mood is used. Our approach can account for this observation, because in this case the formula itself changes the expectations.

**6.4.33. DEFINITION.** (The meaning of the indicative mood)

Let  $M$  be a model,  $g$  an assignment function,  $c$  a cognitive state, and  $\psi$  an expression of type  $[]$ .

$$\begin{aligned} c[IND(\psi)]_{M,g}^+ &= \begin{cases} c[\psi]_{M,g}^+ \text{ if } \forall p \in (c[\psi]_{M,g}^+)_{\eta(c[\psi]_{M,g}^+)} \exists p' \in (c[\psi]_{M,g}^+)_0 : p' \rightsquigarrow p^{72} \\ \text{undefined otherwise.} \end{cases} \\ c[IND(\psi)]_{M,g}^- &= \begin{cases} c[\psi]_{M,g}^- \text{ if } \forall p \in (c[\psi]_{M,g}^-)_{\eta(c[\psi]_{M,g}^-)} \exists p' \in (c[\psi]_{M,g}^-)_0 : p' \rightsquigarrow p \\ \text{undefined otherwise.} \end{cases} \end{aligned}$$

The subjunctive, on the contrary, demands that the update is not everywhere consistent with the expectations of  $c_0$ .

**6.4.34. DEFINITION.** (The meaning of the subjunctive mood)

Let  $M$  be a model,  $g$  an assignment function,  $c$  a cognitive state, and  $\psi$  an expression of type  $[]$ .

$$\begin{aligned} c[SUBJ(\psi)]_{M,g}^+ &= \begin{cases} c[\psi]_{M,g}^+ \text{ if } \exists p \in (c[\psi]_{M,g}^+)_{\eta(c[\psi]_{M,g}^+)} \forall p' \in (c[\psi]_{M,g}^+)_0 : p' \not\rightsquigarrow p \\ \text{undefined otherwise} \end{cases} \\ c[SUBJ(\psi)]_{M,g}^- &= \begin{cases} c[\psi]_{M,g}^- \text{ if } \exists p \in (c[\psi]_{M,g}^-)_{\eta(c[\psi]_{M,g}^-)} \forall p' \in (c[\psi]_{M,g}^-)_0 : p' \not\rightsquigarrow p \\ \text{undefined otherwise} \end{cases} \end{aligned}$$

To motivate the approach to the semantic of the indicative and the subjunctive mood proposed here, let us discuss some of the observations about English this approach can account for. One very intriguing observation is that it is possible to make a statement about the actual world with a subjunctive sentence. This is, in particular, possible for *might* statements.

(138) It might be raining tomorrow.

To account for these sentences is one of the main challenges for approaches to the moods. One might, of course, propose that there is no (semantic) subjunctive in these sentences, but then one has problems accounting for the observation that

there are contexts that demand the subjunctive for other verbs and where *might* has to be used in place of *may*. This is, for instance, the case for counterfactual conditionals and other statements about non-actual possibilities like those given in example (133a) and (133b), here repeated as (139a) and (139b).

- (139) a. He may<sup>?</sup>/might have won the game, but in the end he didn't.  
 b. It may<sup>?</sup>/might have rained this day. Fortunately, it didn't.

This leaves the possibility that the subjunctive is syntactically active in *might*, but not semantically, and that there are syntactic constraints on mood in these contexts. But, intuitively, these contexts often allow a semantic characterization. So, one should first see, whether one can explain the distribution of the subjunctive as due to the meaning of the mood. The present approach proposes such an explanation. The logical form the approach assigns to sentence (138) is  $SUBJ(PRES_m(MOLL_n(P)))$ . This sentence of  $\mathcal{L}$  can give rise to a non-trivial update to a cognitive state  $c$  defined only for index 0. If defined, then the update with the formula  $MOLL_n(P)(d_m)$  introduces a new subordinate context containing the result of  $AIntervene_{M,g}^+(c, P(d_n))$ . This operation copies to the new basic state all possibilities in  $c_0$  where  $P(d_n)$  is defined to be true, and for every possibility  $p$  where  $P(d_n)$  is undefined it will add all possibilities that fill up the future of  $p$  to the point that the truth of  $P(d_n)$  gets defined. But  $AIntervene$  does not consider whether  $P(d_n)$  actually conforms to the expectations in such a possibility  $p$ . That means that also for possibilities  $p$  in  $c_0$  where  $P(d_m)$  is possible, but goes against the expectations,  $AIntervene_{M,g}^+(p, P(d_n))$  will be nonempty. However, none of the elements  $p'$  of this set will follow the expectations of  $p$ . If there is also no other element  $p''$  of  $c_0$  such that  $p'' \rightsquigarrow p'$ , then the indicative mood cannot be used and the subjunctive mood has to be used instead.

Another puzzling observation about subjunctive sentences that has to be explained is that simple subjunctive sentence or subjunctive sentences with the modal *WOLL* cannot be uttered as a statement about the actual world, i.e. be updated to  $c_0$ , but appear to need to refer to some previously introduced hypothetical context.<sup>73</sup> Let us first explain why this is predicted by our approach for

---

<sup>73</sup>There are exceptions in contexts where politeness plays an important role, as in the case of a waitress, responding to the question of one of her guest 'Who is the man over there?'.

- (140) That would be Mr. Smith.

One way to deal with such examples is to take the condition of the subjunctive out of semantics and put it into pragmatics. As has been pointed out by Asher & McCready (2007), another exception are examples of the following form:

- (141) A: Kim teased Pat.  
 B: Kim would do that.

In this case pragmatics does not seem to provide an explanation. Actually, we make in this case the predictions that Asher & McCready think proper for such examples: a cognitive state



simple sentences like (142) with a logical form  $SUBJ(PRES_n(P))$ .

(142) Mary drank all the beer.

Such sentences do not introduce new subordinate contexts. Thus, the condition of the subjunctive mood comes down to the following, where both quantifications run over the same context  $ALearn_{M,g}^+(c, Pres_n(P))_0$ .

$$\exists p \in ALearn_{M,g}^+(c, PRES_n(P))_0 \forall p' \in ALearn_{M,g}^+(c, PRES_n(P))_0 : p' \not\rightsquigarrow p$$

Now, take  $p' = p$ . In this case it certainly holds  $p' \rightsquigarrow p$ . Hence the condition of the subjunctive cannot be fulfilled. Thus, we predict that simple sentences with the subjunctive cannot be about the common ground or the actual world.<sup>74</sup> We come now to the more complicated case of sentences like (143) as an update to  $c_0$ .

(143) Mary would drink all the beer.

Without loss of generality let us consider the ontic reading of the modal. Successful updates with sentences  $SUBJ(PRES_n(WOLL_m(P)))$  do introduce new hypothetical contexts. In this case the condition of the subjunctive comes down to the following (in case the definedness conditions of the present tense and the modal are fulfilled as well as the test condition of the modal).

$$\exists p \in AIntervene_{M,g}^+(c[1], P(d_m))_1 \forall p' \in c_0 : p' \not\rightsquigarrow p$$

On the other hand, in case the update with the subsentence  $WOLL_m(P)(d_n)$  was successful, we know that  $c \models P(d_m)$ , in particular, that the following holds:

$$\forall p \in c_0 \exists p' \in AIntervene_{M,g}^+(c[1], P(d_m))_1 : p \rightsquigarrow p'.$$

With fact 6.4.22 we can conclude:

$$\forall p' \in c_0 \forall p' \in AIntervene_{M,g}^+(c[1], P(d_m))_1 : p \rightsquigarrow p'.$$

---

supports the second sentence in case it also supports the first. However, this does not seem to fit exactly the intuitions about this dialogue. B's utterance also seems to be supported by a cognitive state that does not support that Kim teased Pat. We leave this issue for future work.

<sup>74</sup>A possible problem for the present approach may be that this theory predicts that the subjunctive can in general be used to refer to subordinated contexts – even without being accompanied by a modal. While this is correct for languages like German, in English a modal is needed. A possible explanation is that this obligation has been introduced to prevent confusion of present subjunctive and past indicative readings that without presence of a modal cannot be distinguished. In German, such a confusion cannot arise because of the existence of an independent subjunctive marker.

Hence, the condition of the subjunctive cannot be fulfilled. The crucial aspect of the approach that is responsible for this prediction is that *WOLL* by itself cannot introduce a counterfactual subordinate context. It even cannot introduce violations of the expectations.

Let us also discuss another observation concerning the distribution of sentences with the modal *would*. This is the observation that the use of a *would* conditional is fine even in cases where the consequent is actually true in  $c_0$ . Examples are *even if* conditionals.

- (144) I won't marry you. Even if you had all the money in the world, I would still not marry you.

Such cases could have been a problem for our approach, if we had proposed that the subjunctive mood requires that the sentence in its scope itself is not consistent with the expectations in  $c_0$ . Instead, we demand for the subjunctive mood to be satisfied that the result after updating the consequence has to have this property. Because the consequent of conditionals is updated to the context resulting from update with the antecedent, it is sufficient that the antecedent goes against the expectations of  $c_0$  to make the use of the subjunctive on the consequent acceptable.

On the other hand, we do predict that the subjunctive mood cannot be used if the antecedent is known to be true. This is contra to a claim made by Karttunen & Peters (1977), who defend that *would* conditionals in such circumstances are wellformed. As evidence they provide example (145).

- (145) If Shakespeare were the author of Macbeth, there would be proof in the records of the Globe Theater for the year 1583. So we had better go through them again more carefully until we find that proof.

But this is not a very appropriate example to support their claim. In this example it is not explicitly made clear that at the moment the conditional is stated indeed the antecedent is known to be true. The next less ambiguous example (146a), on the contrary, is unacceptable.

- (146) a. Peter is in the building. And if he \*were/\*was in the building, he would hear us.  
b. Peter is in the building. And if he is in the building, he will hear us.

Also other authors have argued against the standpoint of Karttunen & Peters (1977). The next contrast stems from Iatridou (2000).

- (147) a. If John comes to the party, and I think he will, we will have a great time.

- b. \*If John came to the party, and I think he will, we would have a great time.

So much for the meaning of the indicative and the subjunctive mood. As announced above, we propose that in English there exists also a third mood: the counterfactual mood. The counterfactual mood cancels a basic assumption underlying standard update: the assumption that the update is consistent with what is believed to be the case in the actual world, i.e. consistent with the basic state  $c_0$  of the cognitive state that is updated.<sup>75</sup> The counterfactual mood conveys that the result of the update is not consistent with  $c_0$  (after update).

**6.4.35. DEFINITION.** (The meaning of the counterfactual mood)

Let  $M$  be a model,  $g$  an assignment function,  $c$  a cognitive state, and  $\psi$  an expression of type  $[]$ .

$$\begin{aligned} c[COUNT(\psi)]_{M,g}^+ &= \begin{cases} c[\psi]_{M,g}^+ \text{ if } \forall p \in (c[\psi]_{M,g}^+)_{n(c[\psi]_{M,g}^+)} \forall p' \in (c[\psi]_{M,g}^+)_0 : w_{p'} \not\subseteq w_p \\ \text{undefined otherwise,} \end{cases} \\ c[COUNT(\psi)]_{M,g}^- &= \begin{cases} c[\psi]_{M,g}^- \text{ if } \forall p \in (c[\psi]_{M,g}^-)_{n(c[\psi]_{M,g}^-)} \forall p' \in (c[\psi]_{M,g}^-)_0 : w_{p'} \not\subseteq w_p \\ \text{undefined otherwise.} \end{cases} \end{aligned}$$

According to this approach, counterfactuality is a semantic property of certain English sentences. As the lexicon shows (see figure 6.9 on page 206), these are sentences where the finite verb stands in the simple past and is combined with a perfect. In this case the simple past and the perfect together can be interpreted as conveying the counterfactual mood. This makes it clear that for conditionals only *would have* conditionals can semantically imply counterfactuality. However, we do not predict that all *would have* conditionals convey counterfactuality by their semantics. The perfect marking can be interpreted as semantic perfect, and likewise the simple past can be interpreted as semantic past tense. In particular, the lexicon given in figure 6.9 predicts that if the *would have* conditionals is about the past, then the perfective form has to be responsible for this past shift and, thus, counterfactuality is not a semantic property of the conditional. If the *would have* conditional is about the present or the future, then the perfect together with the simple past occurring in *would have* conditionals is (normally) interpreted as counterfactual mood and the conditional implies counterfactuality semantically. We leave it open whether *would have* conditionals where the perfect is interpreted as semantic perfect can nevertheless pragmatically imply counterfactuality. The pragmatics of conditionals is not our concern here.

Let us finally add some comments on how this proposal of a counterfactual mood relates to the position of other authors on the role of counterfactuality in the meaning of conditionals. There is a lot of discussion in the literature

---

<sup>75</sup>This is the assertion condition we build into the semantics of the update functions *ALearn* and *AIntervene*.

on the relation between subjunctive conditionals and counterfactuality. Some authors claim that all subjunctive conditionals convey counterfactuality (see, for instance, Lewis 1973). Others argue that neither *would* conditionals nor *would have* conditionals carry counterfactuality as part of their semantic meaning. To underpin their position, proponents of this thesis (like, for instance, Karttunen & Peters 1977 and Comrie 1986) often use the following type of example.<sup>76</sup>

- (148) a. If Mary was allergic to penicillin, she would have exactly the symptoms she is showing. (*would* conditional, from Karttunen & Peters 1977)
- b. If the butler had done it, we would have found just the clues that we did in fact find. (*would have* conditional, from Comrie 1986)

Another argument that has been brought forward to support the claim that counterfactuality is only a pragmatic implicature of subjunctive conditionals is that you can assert the falsity of the antecedent afterwards without producing redundancy (see, for instance, Iatridou 2000: 232). These arguments are not particularly strong. The redundancy test is known to be problematic. There are, for instance, also certain types of presuppositions that can be stated without producing redundancy. As to the famous penicillin example, notice that it is about a very particular type of conditional: in the consequent a hypothetical state derived from the antecedent is compared with what is known about the actual world. We have already observed in section 6.2.2 that for such comparisons English uses the subjunctive mood (see example (107b), here repeated as example (149)). Thus, the use of the subjunctive in the penicillin examples may be demanded by the comparison instead of the conditional.

- (149) He behaves like he was sick.

To further support this idea, notice that the penicillin example cannot be restated as indicative conditional.<sup>77</sup>

- (150) \*If Mary is allergic to penicillin, she will show exactly the symptoms she is showing.

If counterfactuality is in general a pragmatic inference, it should be possible to find many more examples where this implicature is cancelled than just the penicillin cases. Indeed, it is easy to find such examples for *would* conditionals. The next sentence, for instance, does not exclude the possibility that I do win the lottery. In general, *would* conditionals about the future are not understood to be counterfactual.

---

<sup>76</sup>According to Ippolito (2003), this type of examples has been introduced by Anderson (1951). While the *would* conditional (148a) is generally accepted, native speakers tend to have different intuitions concerning the acceptability of the *would have* variant (148b).

<sup>77</sup>We are here interested in the reading where antecedent and consequent are about Mary's state at the utterance time, not at some future time. For the future reading the indicative can be used without problems.

(151) If I won the lottery, I would buy a car.

The facts are somewhat less clear for *would have* conditionals. Here a difference has to be made between *would have* conditionals the consequent of which refers to the past and *would have* conditionals the consequent of which refers to the present or the future. Various authors have noticed that in the second case cancellation of counterfactuality is not possible (see, for instance, Dudman 1984, Leirbukt 1991, Ippolito 2003). Examples like (152a) are judged to be semantically anomalous by speakers of English. Ippolito additionally supports this generalization by observing that the penicillin-example cannot be translated into an example about the future (Ippolito 2003: 147, here repeated as (152b)).

- (152) a. \*I'm not sure whether Peter will pay back his debts tomorrow. I doubt it. But if he had paid them back, Mary would have had a lot of money to spend.
- b. \*If Charlie had gone to Boston by train tomorrow, Lucy would have found in his pockets the ticket that she in fact found. So he must be going to Boston tomorrow.

We conclude that in case a *would have* conditional is about the future or the present, then counterfactuality is an obligatory inference of the construction. This confirms our prediction that counterfactuality is a semantic property of *would have* conditionals about the present or future. But what about *would have* conditionals about the past? Here, it seems to be the case that while a counterfactual interpretation is preferred, exceptions are possible. Leirbukt (1991) gives an example from a German newspaper for the German equivalent of a *would have* conditional.

(context: Karl Kaiser, director of the research center of the German Society for foreign politics, claimed in a speech that in a secret declaration of the German contract Adenauer accepted the Oder-Neisse-line as final border to Poland for the case of a peace contract): [...] Bonn ist aufgeschäuscht. Falls sich Karl Kaisers Angaben bestätigen, wäre der heftige innenpolitische Streit um das "Fortbestehen des Deutschen Reiches in den Grenzen von 1937 unabhängig von allen in der Zwischenzeit geschlossenen Verträgen", der inzwischen sogar an die Substanz der Bonner Koalition geht, absolut gegenstandslos. Mit einer Verzichtserklärung hätte Adenauer alle Nachfolgeregerungen rechtsverbindlich gebunden (Weser-Kurier 137.1989, S.2, markings are added by the author).

Similar cases for English seem possible as well. Suppose that for a school anniversary someone asks you whether Mary got married or whether she still carries her old name. In this context many speakers of English agree that it is acceptable to say the following.

- (153) I don't know whether Mary got married. When I knew her, she always told me that she despises marriage. But if – despite of her former aversions – she had married, she wouldn't have changed her name.

Furthermore, even native speakers that have problems with constructions like (153) admit that sentences like (153), and (154a) and (154b) given below, are much more acceptable than non-counterfactual *would have* conditionals with future evaluation time like (152a).

- (154) a. I don't know whether Peter won the race yesterday. But if he had, he would have been very happy.  
 b. I'm not sure whether Peter paid his debts yesterday. But if he had paid them back, Mary would have had a lot to celebrate.

These observations imply that counterfactuality cannot be part of the semantics for all *would have* conditionals referring to the past. However, they leave the question open, whether this is true for some of these conditionals. There is some evidence to the point that this might be the case. At the same time this evidence also supports our claim that it is the perfect that encodes the counterfactuality semantically. There is apparently a historical process of change going on in English, in British English and even stronger in American English. Native speakers show an growing tendency to group in *would have* constructions the auxiliary *have* together with the modal and not with the past participle. This process seems to move towards a new *would have* conditional where *have* has developed into a suffix of the modal. This is supported by a corpus study of Boyland (1998), where the author shows that in earlier stages of English adverbials and parentetical expressions rarely occurred between the auxiliary *have* and the past participle in *would have* constructions, but now such constructions become more and more frequent, particularly in spoken English. Boyland also mentions a number of other observations showing the growing inseparability of *would* and *have* that seems to lead towards *have* becoming affixed onto *would*. Our approach could explain these observations as the result of a change of meaning of the perfect auxiliary to a marker of the counterfactual mood. Mood is in general assumed to scope semantically over the modal. This may explain the tendency to transform the auxiliary into a suffix. A related observation noticed by various authors is that in the antecedent of *would have* conditionals (and *would have* constructions in general) the auxiliary *have* is often doubled in occurrence (the data are corpus-results and except for the last taken from Boyland 1998).

- (155) a. But do you think the QCs would have still have linked ...?  
 b. They claimed they did the best they could have have, ... .  
 c. I would have had done a ten times better job if ... .

- d. ... heat ... could have had melted the crusts of both moons
- e. ... something I've never would have done.
- f. There is no need to come in and in fact if anybody had've done, they'd have been told to get out.

Boyland comments: "Such multiplication of forms could mean that the functions of *have*, in modal perfect constructions, are splitting up and become distinct from one another (Denison, 1993). If this is the case, the functional split may be being driven by the grammaticalization of *would've*, which appropriates one *have* for its own counterfactual purposes, leaving the past-marking function to be filled by a second *have*." (Boyland 1998). The present approach can explain the multiple occurrences of *have* as due to the need for using the semantic perfect to make a distinction between counterfactual *would have* conditionals about the past and counterfactual *would have* conditionals about present and future.

These observations can be taken to argue that also in *would have* conditionals referring to the past the past perfect can semantically encode the counterfactual mood. In *would have* conditionals with deviating syntax the auxiliary *have* does not seem to express the perfect meaning anymore. But because the syntactic deviations can be observed for *would have* conditionals about the present and the future, as well as for *would have* conditionals about the past the difference in meaning appears to apply equally to all *would have* conditionals. We have argued that the past perfect can semantically convey counterfactuality. Mood in English is generally marked in the morphology of the finite verb. Thus, one might propose that the apparent change in function of the auxiliary is one to a mood marker. But then this change is not restricted to *would have* conditionals about the present and the future.

We suggest the following explanation for the data. English is in a stage where the auxiliary *have* is changing its function in *would have* conditionals. It develops from a marker of the semantic perfect into an affix expressing the counterfactual mood. At the moment still both interpretations are available. In case the interpretation is the semantic perfect, then the *would have* conditional refers (normally) to the past and counterfactuality is not part of the semantic meaning of the conditional. If *have* is interpreted (together with the past tense marker) as the counterfactual mood, then the conditional is counterfactual by semantics. In this case there is no part of the construction that restricts the evaluation time of antecedent and consequent. They can be located at any time in the past, present, or future.

To build the thesis that *would have* conditionals about the past can also semantically convey counterfactuality into our framework, we have to make some changes in our lexicon (see figure 6.12 on page 246). We introduce a new *ZERO* tense that is interpreted as a temporal variable without restrictions on possible referents. Furthermore, we do not let the finite verbform, but the mood, select

a semantic tense: the subjunctive mood asks for the present tense, while the counterfactual mood asks for zero tense.



A VARIATION OF THE LEXICONCategory: PROPERTY

semantic expression	type	syntactic features	realization
<i>P</i>	[i]	[-ind, -pres]	<i>Mary-drinks-all-the-wine</i>
<i>P</i>	[i]	[-ind, -past]	<i>Mary-drank-all-the-wine</i>
<i>P</i>	[i]	[-subj]	<i>Mary-drank-all-the-wine</i>
<i>P</i>	[i]	[-perf]	<i>Mary-drunk-all-the-wine</i>
<i>P</i>	[i]	[]	<i>Mary-drunk-all-the-wine</i>
<i>P</i>	[i]	[]	<i>Mary-drink-all-the-wine</i>
...	...	...	...

Category: MODAL

semantic expression	type	syntactic features	realization
<i>WOLL<sub>n</sub></i>	[[i]i]	[-ind, -pres]	<i>will</i>
<i>WOLL<sub>n</sub></i>	[[i]i]	[-ind, -past]	<i>would</i>
<i>WOLL<sub>n</sub></i>	[[i]i]	[-subj]	<i>would</i>
<i>MOLL<sub>n</sub></i>	[[i]i]	[-ind, -pres]	<i>may</i>
<i>MOLL<sub>n</sub></i>	[[i]i]	[-subj]	<i>might</i>

Category: ASPECT

semantic expression	type	syntactic features	realization
<i>PERF<sub>n</sub></i>	[[i]i]	[-ind, -pres, +perf]	<i>have</i>
<i>PERF<sub>n</sub></i>	[[i]i]	[+perf]	<i>have</i>
	[[i]i]	[-count]	<i>have</i>
<i>PERF<sub>n</sub></i>	[[i]i]	[-ind, -past, +perf]	<i>had</i>
<i>PERF<sub>n</sub></i>	[[i]i]	[-subj, +perf]	<i>had</i>
	[[i]i]	[-count, -subj]	<i>had</i>

Category: TENSE

semantic expression	type	syntactic features	realization
<i>PRES<sub>n</sub></i>	[[i]]	[+pres]	*
<i>PAST<sub>n</sub></i>	[[i]]	[+past]	*
<i>ZERO<sub>n</sub></i>	[[i]]	[+zero]	*

Category: MOOD

semantic expression	type	syntactic features	realization
<i>IND</i>	[[ ]]	[+ind]	*
<i>SUBJ</i>	[[ ]]	[+subj, -pres]	*
<i>COUNT</i>	[[ ]]	[+count, +subj, -zero]	*

Category: CONNECTIVES

semantic expression	type	syntactic features	realization
<i>IF</i>	[[ ] [ ]]	[]	<i>if</i>
¬	[[ ]]	[]	<i>not</i>
∧	[[ ] [ ]]	[]	<i>and</i>
∨	[[ ] [ ]]	[]	<i>or</i>

Figure 6.12: The adapted lexicon of our fragment of English

#### 6.4.4.9 The meaning of $IF$

The last ingredient we need before we can give a compositional derivation of the meaning of conditionals is a semantics for *if*. We propose that *if*, like *and* and *or*, is a two-place sentential operator. As for the modals, the main semantic contribution of *if* lies in performing a test on the cognitive state that is updated with the conditional. If the test is successful, the update results in the introduction of a new subordinate context to which both, antecedent and consequent contribute their meaning. Following the previous chapter, we distinguish two readings for *if*: an ontic reading ( $IF_1$ ) and an epistemic reading ( $IF_2$ ). The difference between the two operators  $IF_1$  and  $IF_2$  lies in whether they refer to the epistemic or the ontic update function in the test. The epistemic reading of *if* tests whether a hypothetical *epistemic* update with the antecedent would *support* the consequent. The ontic reading of *if* tests whether a hypothetical *ontic* update with the antecedent would *force* the consequent. In both cases the restriction of the update functions to consistent updates is lifted. Thus, in the context of conditionals the update functions may truly revise a basic state. The definition of the epistemic and the ontic update functions without the restriction to consistent updates are given following the interpretation rules for the two readings of *if*. These definitions build on the theory developed in Chapter 5.

##### 6.4.36. DEFINITION. (The interpretation rules for $IF_1$ and $IF_2$ )

Let  $M$  be a model,  $g$  and assignment function,  $c$  a cognitive state, and  $\psi$  and  $\phi$  expressions of type  $\square$ .

$$\begin{aligned} c[IF_1 \psi, \phi]_{M,g} &= \begin{cases} Intervene_{M,g}^+(c[\eta(c)], \psi \wedge \phi) \text{ if } Intervene_{M,g}^+(c[\eta(c)], \psi) \models \phi, \\ c[\eta(c)/\emptyset] \text{ if } Intervene_{M,g}^+(c[\eta(c)], \psi) \not\models \phi, \end{cases} \\ c[IF_2 \psi, \phi]_{M,g} &= \begin{cases} Learn_{M,g}^+(c[\eta(c)], \psi \wedge \phi) \text{ if } Learn_{M,g}^+(c[\eta(c)], \psi) \models \phi, \\ c[\eta(c)/\emptyset] \text{ if } Learn_{M,g}^+(c[\eta(c)], \psi) \not\models \phi. \end{cases} \end{aligned}$$

**Extending  $ALearn$  to the revision case.** In this section we will extend the update function  $ALearn$  so that it can also deal with counterfactual updates. We want this extension to take the same stance towards belief revision as does the description of the epistemic reading of *would have* conditionals provided in Chapter 5. It is not straightforward to extend the approach made there to the dynamic and time-sensible framework we are working with in this chapter. One problem is that the approach to belief revision introduced in Chapter 5 works on the basis of the set of facts for which the agent of some epistemic state has independent external evidence. Somehow, we have to build this information into our formal notion of a cognitive state. Another problem is that the definition of the function  $Learn$  provided in Chapter 5 does not describe this function in a strict sense as one of belief revision: it does not return a belief state. In consequence, we cannot iterate applications of  $Learn$ . This was not a problem for the use we made of  $Learn$  in Chapter 5, because we only applied it for the evaluation of *would have*

conditionals and we did not allow for iterated uses of the conditional operator. It is also not a direct problem of the formal setup of the present framework, because the same restrictions hold in principle here as well (*Learn* only occurs in the interpretation rule for *IF* and iterated occurrences of the conditional could be suppressed). But remember that the conceptual idea behind the distinction between the functions *ALearn* and *Learn* introduced in this chapter is that *ALearn* is what you get if you add to *Learn* the (pragmatic) well-formedness condition of assertions that the update is consistent. That means that once we implement a pragmatic theory that accounts for this well-formedness condition of assertions, we should be able to do without the function *ALearn* and always apply *Learn*. But in order to be able to substitute *Learn* for all occurrences of *ALearn* in the interpretation rules, we would need *Learn* to be a true function of belief revision. Otherwise, it could not serve as general update rule. We will come back to this problem at the end of this section.

In the following we make the simplifying assumption that (epistemic) conditionals always refer to  $c_0$ , the basic state that represents the information available about the actual world. That means that we do not consider conditional claims about hypothetical contexts. The second assumption we make is that the facts the agent of some cognitive states has external evidence for are given by the set of sentences updated to the cognitive state. In fact, it is not enough to simply memorize all incoming sentences. We also need the order in which they were added to these cognitive states in order to keep the correct dynamic relations between them.<sup>78</sup> We will keep track of the facts the agent has external evidence for by extending a cognitive state with a set  $\mathcal{B}$  of tuples consisting of a sentence of  $\mathcal{L}$  and a natural number. The sentences are those sentences updated to the cognitive state and the numbers encode the order in which they were updated.

**6.4.37. DEFINITION.** (An extended notion of a cognitive state)

Let  $M$  be a model and  $g$  an assignment function.  $c^T$  is the (unextended) cognitive state only defined for index 0 where  $c_0^T$  is the basic state containing all possibilities of model  $M$  that obey the laws of the law structure in  $M$ . An *extended cognitive state* is a tuple  $\langle \mathcal{B}, c \rangle$ , where  $\mathcal{B}$  is a set of tuples  $\langle \varphi, i \rangle$  with  $\varphi \in \mathcal{L}$  and  $i$  a natural number (every number can only occur once),  $c$  is a (non-extended) cognitive state, and  $ALearn_{M,g}^+(c^T, \mathcal{B}) = c$ .  $ALearn_{M,g}^+(c^T, \mathcal{B})$  is the subsequent update of the sentences in  $\mathcal{B}$  to  $c$ , following the order of the numbers the sentences are associated with. We call  $\mathcal{B}$  the *basis* of the extended cognitive state  $\langle \mathcal{B}, c \rangle$ .

There is one part of this definition that still has to be explained. We defined  $c_0^T$  as the set of possibilities of  $M$  and  $g$  that obey the laws of  $M$ . In general,

---

<sup>78</sup>We should also store together with the sentences all the contextual parameters needed to resolve deictic reference. However, this last aspect we will ignore to make matters not too complicated.

we do not demand that a possibility has to obey all laws of  $M$ . Instead we require that a possibility does not violate the analytical/logical laws of  $M$  (see the definitions 6.4.9 and 6.4.14). We need possibilities that violate causal laws for the ontic update function. They are, for instance, essential for our explanation of why causal backtracking is not possible for ontic conditionals about the past or the present. But we also argued in Chapter 5 that the epistemic reading of conditionals allows for causal backtracking. To model this, we have to demand that belief revision is calculated with respect to the smaller set of possibilities that also obey the causal laws. We define what it means for a possibility  $p$  to obey the causal laws below. We will say in this case that  $p$  is *faithful* to the law structure  $\langle C, U \rangle$  of a model  $M$ .

**6.4.38. DEFINITION.** (Faithfulness)

Let  $M = \langle S, \langle T, < \rangle, \langle C, U \rangle, I, now \rangle$  be a model for the language  $\mathcal{L}$  with  $C = \langle B, E, F \rangle$  and  $g$  an assignment function. An interpretation function  $w : (\mathcal{L} \times T) \longrightarrow \{0, 1\}$  is *faithful to the law structure*  $\langle C, U \rangle$  if it satisfies the following conditions:

- (i)  $w \in U$ ,
- (ii) for all  $P \in E$  with  $Z_P = \langle P_1, \dots, P_n \rangle$  and all  $i \in I(T)$ ,  $\overline{w}(P, i)$  is defined and there exists truth values  $x_1, \dots, x_n$  such that  $f_P(x_1, \dots, x_n) = \overline{w}(P, i)$  if and only if there exists  $j \in I(T)$  such that  $j < i$  &  $\neg \exists t \in T(j < t < i)$  and for all  $k \in \{1, \dots, n\}$ ,  $\overline{w}(P_k, j)$  is defined,  $f_P(\overline{w}(P_1, j), \dots, \overline{w}(P_n, j))$  is defined, and  $f_P(\overline{w}(P_1, j), \dots, \overline{w}(P_n, j)) = \overline{w}(P, i)$ .

A possibility  $p \in W_{M,g}$  is *faithful to the law structure*  $\langle C, U \rangle$  if there exists some interpretation function  $w$  faithful to the law structure  $\langle C, U \rangle$  such that  $w_p \subseteq w$ .

Now, we come to the definition of belief revision on the level of cognitive states. Let us start with the introduction of some useful notation. Let  $\mathcal{B}$  be a set of tuples  $\langle \varphi, i \rangle$  with  $\varphi \in \mathcal{L}$  and  $i$  a natural number (every number can occur only once), and  $\psi$  be a sentence of  $\mathcal{L}$ . Then  $\mathcal{B} + \psi$  denotes the extension of  $\mathcal{B}$  with a tuple  $\langle \psi, n \rangle$  where  $n \in \mathbb{N}$  is a successor of the maximal  $k \in \mathbb{N}$  with  $\langle \phi, k \rangle \in \mathcal{B}$ . Furthermore, we say that  $\mathcal{B}$  is *satisfiable* if  $ALearn_{M,g}^+(c^T, \mathcal{B}) \notin \perp$ . The informal idea we apply here when modeling belief revision is exactly the same as that underlying the formalization of the function *Learn* provided in Chapter 5. Belief revision tries to keep all laws and as many as possible of the basis facts of a cognitive state. But we cannot just take as output of revising  $\langle \mathcal{B}, c \rangle$  with  $\psi$  a cognitive state with a basis  $\mathcal{B}' + \psi$ , where  $\mathcal{B}'$  is a maximal subset of  $\mathcal{B}$  such that  $\mathcal{B}' + \psi$  is satisfiable. The reason is that there can be more than one such maximal subset. We will apply the same solution for this problem employed in Chapter 5 and take as output of belief revision the union of all cognitive states

with a basis of the form  $\mathcal{B}' + \psi$ , where  $\mathcal{B}'$  is a maximal subset of  $\mathcal{B}$  such that  $\mathcal{B}' + \psi$  is satisfiable.

**6.4.39. DEFINITION.** (Dynamic belief revision)

Let  $M$  be a model,  $g$  an assignment function,  $\langle \mathcal{B}, c \rangle$  an extended cognitive state, and  $\psi$  a sentence of  $\mathcal{L}$ . We define

$$[\mathcal{B}] = \{ \mathcal{B}' \subseteq \mathcal{B} \mid \begin{aligned} &ALearn_{M,g}^+(c^T, \mathcal{B}' + \psi) \notin \perp \text{ \& } \\ &\neg \exists \mathcal{B}'' \subseteq \mathcal{B} : ALearn_{M,g}^+(c^T, \mathcal{B}'' + \psi) \notin \perp \text{ \& } \mathcal{B}'' \supset \mathcal{B}' \end{aligned} \}.$$

The epistemic update of an extended cognitive state  $\langle \mathcal{B}, c \rangle$  with the sentence  $\psi$  is then defined as follows.

$$\begin{aligned} Learn_{M,g}^+(\langle \mathcal{B}, c \rangle, \psi) &= \langle \mathcal{B}, c[\eta(c)/(\bigcup_{\mathcal{B}' \in [\mathcal{B}]} ALearn_{M,g}^+(c^T, \mathcal{B}' + \psi))_0] \rangle \\ Learn_{M,g}^-(\langle \mathcal{B}, c \rangle, \psi) &= \langle \mathcal{B}, c[\eta(c)/(\bigcup_{\mathcal{B}' \in [\mathcal{B}]} ALearn_{M,g}^+(c^T, \mathcal{B}' + \neg\psi))_0] \rangle \end{aligned}$$

The drawback of this definition is – and here we come back to the problem mentioned at the beginning – that the output of *Learn* is not an extended cognitive state. We no longer have in general for  $\langle \mathcal{B}', c' \rangle = Learn_{M,g}^+(\langle \mathcal{B}, c \rangle, \psi)$  that  $c' = ALearn_{M,g}^+(c, \mathcal{B}')$ . Actually,  $\mathcal{B} = \mathcal{B}'$ , thus the revision does not affect the basis. This is a stipulation we make to deal with the problem that it is not clear how to change the basis in case there is more than one maximal satisfiable subset of  $\mathcal{B}$ . Because of the restricted application we make of the function *Learn* (it only occurs in the update rules of conditionals) this is not a problem. As long as we distinguish a rule for consistent update from a rule for inconsistent update, we can define the consistent update *ALearn* as returning an extended cognitive states: we simply extend *ALearn* with the condition that the sentences it is applied to is added to the basis of the extended cognitive state it is applied to. But once we want to do without this distinction of two interpretation rules, we have to reconsider the approach towards belief revision taken here.<sup>79</sup>

**Extending *AI Intervene* to the revision case.** Extending the ontic interpretation function to the revision case is not very difficult. We take the definitions of Chapter 5 for *Intervene* and add those few adaptations we worked out in section 6.4.4.2 to deal with the ontic reading of statements about the future. We start again by defining two orders that describe similarity with respect to bases and derivable facts.

---

<sup>79</sup>There is another complication of updates of extended cognitive states, which we have not mentioned so far. If we apply an interpretation function to an extended cognitive state we have to make a distinction between an update with an independent sentence and an update with some part of this sentence that has to be calculate in order to determine the meaning of the sentence. Only in the first case the basis of the extended cognitive state should be changed.

**6.4.40. DEFINITION.** (The orders for *Intervene*)

Let  $M$  be a model,  $g$  an assignment function, and  $p, p_1, p_2 \in W_{M,g}$ .

$$\begin{aligned}
 p_1 \leq_1^p p_2 \quad \text{iff} \quad & (i) \quad b_{p_1} \cap b_p \supseteq b_{p_2} \cap b_p, \text{ and} \\
 & (ii) \quad \text{if } b_{p_1} \cap b_p = b_{p_2} \cap b_p, \text{ then } (b_{p_1} - B) - b_p \subseteq (b_{p_2} - B) - b_p, \\
 p_1 \leq_2^p p_2 \quad \text{iff} \quad & (i) \quad (w_{p_1} - b_{p_1}) \cap (w_p - b_p) \supseteq (w_{p_2} - b_{p_2}) \cap (w_p - b_p), \text{ and} \\
 & (ii) \quad \text{if } (w_{p_1} - b_{p_1}) \cap (w_p - b_p) = (w_{p_2} - b_{p_2}) \cap (w_p - b_p), \\
 & \quad \text{then } (w_{p_1} - b_{p_1}) - (w_p - b_p) \subseteq (w_{p_2} - b_{p_2}) - (w_p - b_p).
 \end{aligned}$$

The first order,  $\leq_1^p$ , compares similarity with respect to the bases. It first maximizes the overlap with the basis of  $p$  and in a second step minimizes the new miracles introduced. The second order,  $\leq_2^p$ , compares similarity with respect to the derived facts of  $p$ . Again, first the overlap with  $p$  is maximized, and then the difference, the new derivable facts that are introduced, minimized. The differences with the definitions of the orders used in Chapter 5 are first the additional second condition for the order  $\leq_2^p$ . This condition ensures that the possibilities selected by the function *Intervene* do not extend unnecessarily far into the future. We did not need it in Chapter 5, because on the level of abstraction this chapter was working on there was no future. A second difference is that extensions of the basis are only compared with respect to the miracles added. This adaptation we need to allow for causal backtracking in the future. For the rest the orders are identical to those defined in definition 5.6.16 on page 145 of the previous chapter. The definition of *Intervene* even works completely on a par with definition 5.6.17, page 145.

**6.4.41. DEFINITION.** (Intervention for atomic formulas)

Let  $M$  be a model,  $g$  an assignment function,  $p \in W_{M,g}$ ,  $P \in \mathcal{P}$ , and  $d \in VAR_i$ .

$$\begin{aligned}
 Intervene_{M,g}^+(p, P(d)) &= Min(\leq_2^p, Min(\leq_1^p, \llbracket P(d) \rrbracket_{M,p}^+)), \\
 Intervene_{M,g}^-(p, P(d)) &= Min(\leq_2^p, Min(\leq_1^p, \llbracket P(d) \rrbracket_{M,p}^-)).
 \end{aligned}$$

The idea behind the introduction of the operation *AIIntervene* in section 6.4.4.2 was that this function is what we get if we add to the full-blooded version of the ontic interpretation function *Intervene* the condition that assertions have to be consistent with the cognitive state to which they are updated. Thus, we would like to have a result like the following.

Let  $M$  be a model,  $g$  an assignment function,  $p \in W_{M,g}$ ,  $P \in \mathcal{P}$  and  $d \in VAR_i$ . If  $AIIntervene_{M,g}^+(p, P(d)) \neq \emptyset$ , then  $Intervene_{M,g}^+(p, P(d)) = AIIntervene_{M,g}^+(p, P(d))$ . If  $AIIntervene_{M,g}^-(p, P(d)) \neq \emptyset$ , then  $Intervene_{M,g}^-(p, P(d)) = AIIntervene_{M,g}^-(p, P(d))$ .

Unfortunately, in general this does not hold. The central problem is that, depending on which analytical/logical laws we take to be valid, it might not be

true that from  $w_p \subseteq w_{p'}$  we can conclude  $b_p \subseteq b_{p'}$ . Analytical/logical laws that allow you to reason from the future to the past, but not the other way around can destroy this relation. It is very difficult to come up with natural examples for such laws and to evaluate the different predictions made by *AIntervene* and *Intervene* in this point. So far, we do not feel that we are able to make a decision about which of the two functions fares better. We leave this issue to future work.

If, however, one excludes analytical/logical laws and looks on models where only causal laws exists, then the desired relation can be shown to hold.<sup>80</sup> Thus, in this case we can conclude that *Intervene* does the same with atomic formulas that do not involve revision of the facts of the possibility as does *AIntervene*. But even then we still need to get an idea of what *Intervene* actually does in case  $P(d)$  is really counterfactual. This is illustrated with an example. The general setting is the same as that used before. Let  $\mathcal{P}$  be the set of letters containing only  $A$ ,  $B$ , and  $C$ . As time structure  $T$  we take  $\mathbb{Z}$ . Furthermore, we take the law structure  $L = \langle C, U \rangle$  where  $U$  is the set of all complete interpretation functions for  $\mathcal{P}$  and  $T$ , and  $C = \langle B, E, F \rangle$  contains two laws:  $B = \{A\}$ ,  $E = \{B, C\}$ ,  $F(B) = \langle Z_B, f_B \rangle$  with  $Z_B = \langle A \rangle$  and  $f_B = \{\langle 1, 1 \rangle, \langle 0, 0 \rangle\}$ , and  $F(C) = \langle Z_C, f_C \rangle$  with  $Z_C = \langle B \rangle$  and  $f_C = \{\langle 1, 1 \rangle, \langle 0, 0 \rangle\}$ . Let  $p$  be a possibility with the following properties:  $t_p = 0$ ,  $g_p$  is only defined for  $d$ ,  $g(d) = -2$ , and  $w_p$  is the function mapping for all times  $t' \leq 0$   $A$ ,  $B$ , and  $C$  to 0 and is undefined for all other combinations of times and properties.

We want to calculate the result of applying *Intervene*<sup>+</sup> to  $p$  and  $B(d)$ . In figure 6.13 a number of possibilities that the reader may consider potential elements of  $\text{Intervene}_{M,g}^+(p, B(d))$  are described.  $p1$  is the possibility that differs in some sense least from  $p$  and nevertheless makes  $B(d)$  true. The evaluation of  $B$  at  $-2$  is changed, the rest stays the same. This possibility contains two miracles, as Lewis would say: at two times causal laws are broken. First, because  $B$  is true at  $-2$  even though  $A$  is false at  $-3$ , and second because  $C$  is false at  $-1$  even though  $B$  is true at  $-2$ . In the second possibility,  $p2$ , changes have been made to obey the causal consequences of making  $B$  true at  $-2$ . Hence,  $C$  is put to 1 at  $-1$ . The third possibility  $p3$  also adapted the history in a way that  $B$  turns out to be true at  $-2$ . Thus,  $A$  is changed to 1 at  $-3$ . This possibility, together with  $p$ , is the only possibility in the list that obeys all causal laws.  $p4$  closes the world under causal consequence after  $-2$ . This leads to some predetermination of the future.  $p5$  changes  $B$  at  $-2$  to 1 and leaves after this change, everything open. According to the possibility  $p6$ , only the evaluation of the causal consequence  $C$  of  $B$  is undefined after the change in the interpretation of  $B$ . Finally,  $p7$  is like  $p2$  except that it fills in some part of the future as well. We want to know which of these possibilities end up in  $\text{Intervene}_{M,g}^+(p, B(d))$ . The possibility  $p$  is certainly eliminated, because it does not make the formula  $B(d)$  true. Furthermore,  $p4$  and  $p6$  are not in the update, because they are not possibilities according to

---

<sup>80</sup>For a proof see the appendix.

definition 6.4.14. This leaves us with the set  $\{p1, p2, p3, p5, p7\}$  for which we have to calculate how the enclosed possibilities relate with respect to the orders  $\leq_1^p$  and  $\leq_2^p$ . To do this we need the bases of all these possibilities. They are again marked in figure 6.13 by drawing a box around those entries that describe the basis. The graphic on the right side of the figure illustrates how these possibilities relate with respect to the orders. A thick arrow points from possibility  $p_i$  to possibility  $p_j$  if  $p_i <_1^p p_j$ . A thin arrow represents the order  $<_2^p$ . We only mark the relations introduced by the second order for those possibilities minimal with respect to  $\leq_1^p$ .  $p3$  and  $p5$  end up very high in the order, because both change the basis of  $p$ . Changes in the basis have to be prevented with highest priority.  $p1$  is subminimal because it introduces two miracles, while to make  $B(d)$  true only one is needed. Finally,  $p7$  is eliminated because it extends the interpretation function unnecessarily far into the future. Thus  $Intervene_{M,g}^+(p, B(d)) = \{p2\}$ . Given the way possibilities are defined here, it is clear that  $t_{p2}$  has to be 0. Hence, in contrast to the application of  $AIntervene$  and  $Intervene$  to atomic formulas about the future, an application to statements about the past does not result in a shift of the temporal perspective. This accounts for the observation that the puzzle of the shifted temporal perspective does not extend to conditional statements about the past.

Finally, we define *Intervene* for cognitive states. This definition is structurally identical to the definition of *AIntervene* for cognitive states.

**6.4.42. DEFINITION.** Intervention for cognitive states.

Let  $M$  be a model,  $g$  an assignment function,  $c$  a cognitive state,  $P \in \mathcal{P}$ , and  $d \in VAR_i$ . The ontic update of  $c$  with  $P(d)$  is defined as follows:

$$\begin{aligned} Intervene_{M,g}^+(c, P(d)) &= c[\eta(c)/\bigcup_{p \in c_{\eta(c)}} Intervene_{M,g}^+(p, P(d))], \\ Intervene_{M,g}^-(c, P(d)) &= c[\eta(c)/\bigcup_{p \in c_{\eta(c)}} Intervene_{M,g}^-(p, P(d))]. \end{aligned}$$

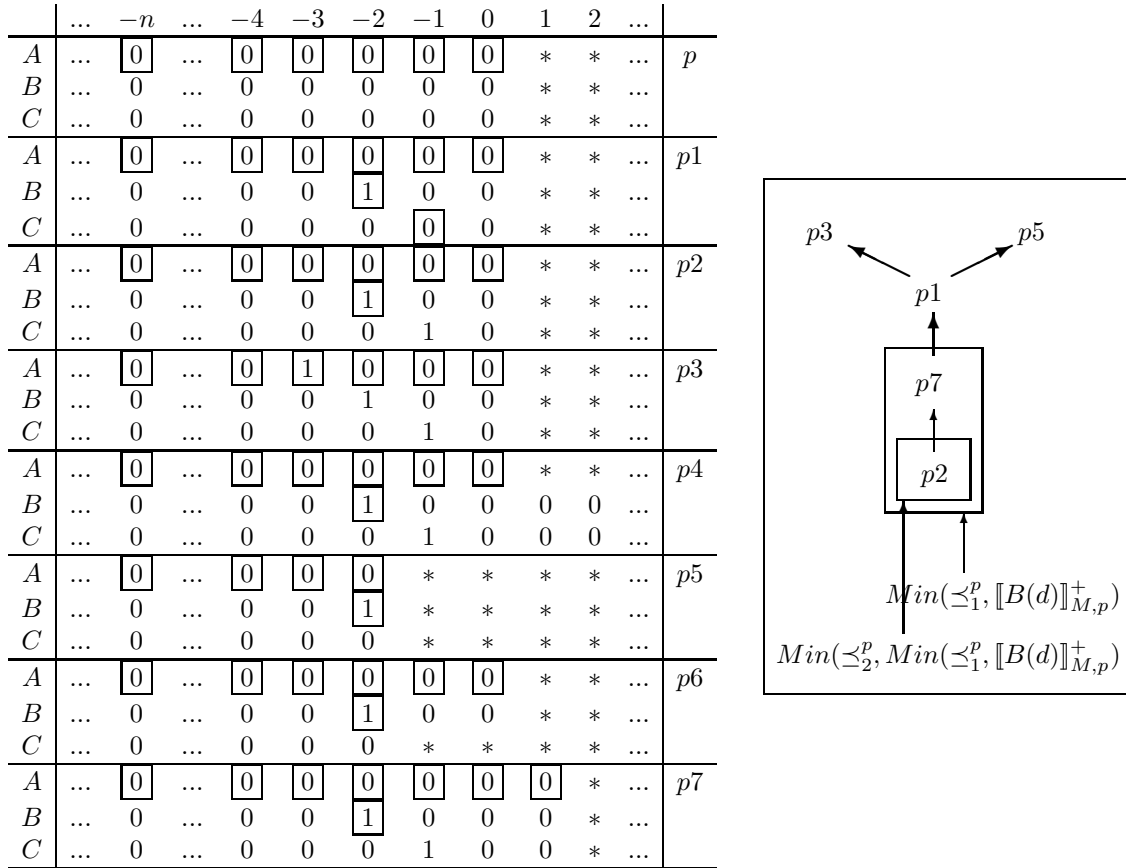
## 6.5 Discussion

In the previous section a compositional approach to the semantic meaning of English conditional sentences has been introduced. It is not a full compositional approach to conditionals, because we did not analyze the syntactic structure of English sentences down to the level of predicate structure. But we did distinguish semantic contributions for the tenses, modals, the perfect and sentential connectives like *if*, *and*, *or*, and *not*, and described how their meaning contributes to the meaning of English conditionals.

At the beginning of this chapter we discussed two puzzles about the interpretation of tense in conditional sentences that we wanted to account for: the puzzle of the missing interpretation and the puzzle of the shifted temporal perspective.



Can the account we proposed in the last section deal with these puzzling observations? The first puzzle, the puzzle of the missing interpretation, concerned the interpretation of the simple past and the perfect in *would* and *would have* conditionals. We observed that in these constructions the expected past-shift of the evaluation time for antecedent and consequent that standard approaches to the meaning of the simple past and the perfect would predict is absent. This is explained by the present approach by distinguishing two semantic meanings for the simple past and the perfect: a temporal meaning and a mood meaning. The temporal meaning follows in both cases standard lines. Besides this temporal meaning the simple past can also be interpreted as expressing the subjunctive mood. The subjunctive mood demands the outcome of the update with the formula in its scope to be inconsistent with the expectations about the actual world of the cognitive state the sentence is updated to. The perfect can in combination with the past tense be interpreted as counterfactual mood. This mood claims that the update with the formula in the scope of the perfect is inconsistent with

Figure 6.13: Some possibilities for  $B(d)$  and their relation

what is known about the actual world. The basic ideas behind this approach to the mood are not new (see, for instance, Quirk et al. 1985: 1091-1093). What is new is the precise formalization of these ideas provided here. The proposed lexical ambiguity for the syntactic past tense and the syntactic past perfect allows us to account for the puzzle of the missing interpretation of English conditionals. The temporal interpretation is predicted to be missing exactly in those cases where their syntactic expressions are interpreted as mood markers. In the classification of the approaches to the puzzle of the missing interpretation introduced in section 6.2 the present proposal falls under past-as-modal approaches. It has the common disadvantage of all approaches in this class of multiplying meanings in the lexicon. On the other hand, it also shares their advantage: the logical form we assign to conditional sentences stays very close to the surface syntactic structure. No movement of operators is involved. We have also seen that besides the puzzle of the missing interpretation the proposed semantics for the English mood can account for characteristic observations about the distribution of the non-temporal meaning of the simple past and past modals.

The second puzzle discussed at the beginning of this chapter was the puzzle of the shifted temporal perspective. We observed that the antecedent of indicative conditionals can shift the reference time for the interpretation of tenses in the consequent to the future. This is possible in case the evaluation time of the antecedent lies in the future. The reference time of tense in the consequent is in these cases (normally) set to the future evaluation time of the antecedent. Something similar we observed for the interpretation of tenses in relative clauses in the scope of modals. But let us start with discussing how our semantics explains the shift of the temporal perspective in the consequent of conditionals. Building on the work of the previous chapter we distinguish two readings for conditionals: an epistemic reading and an ontic reading. While in Chapter 5 we proposed this ambiguity only for *would have* conditionals this claim is now extended to conditional sentences in general. The formalization of the two readings is based on the proposal of Chapter 5. However, the introduction of time into the model and the more complex formal language made some adaptations to the framework necessary. But these changes are conservative in that we can still account for the observations made in the previous chapter for the meaning of *would have* conditionals. The present approach differs from many other approaches towards the meaning of conditionals and the semantics of modals in that it realizes the ambiguity between an ontic and an epistemic reading not by a contextually given variable for a modal base over which the conditional or the modals quantify, but by distinguishing between two semantic interpretation functions to which *IF* and the modals can refer: an epistemic update function and an ontic update function. The epistemic interpretation function takes the formula a cognitive state is updated with as providing information about how the world looks. In the revision-free case it comes down to standard dynamic update as we know it. This reading predicts no shift in the temporal perspective. The ontic interpretation

function takes the formula that is to be updated as prescription. This function changes the world so that it obeys the content of the formula. If the truth value of the formula is defined to be true, then this function does exactly the same as the epistemic update function and keeps this possibility. If the truth value of the formula is defined to be false, then this possibility is either thrown away (revision-free case, the modals) or the function changes the course of events minimally to make the formula true (revision case, conditionals). In none of these cases is a shift of the temporal perspective of the possibility predicted. If, however, the truth value of the formula is undefined, and, thus, concerns some aspect of the future of  $p$ , then the ontic update function fills up the future of  $p$  minimally such that the truth value of the formula gets defined. In this case the temporal perspective of the possibility is shifted forward to this point in the future where the truth value becomes defined. This explains why present tense antecedents of indicative conditionals can shift the reference time for the interpretation of tense in the consequent forward. It also explains why this shift can be absent if the antecedent is interpreted as predetermined at the utterance time or why sometimes the shift only goes to some time in the future at which the truth becomes predetermined. The ontic reading only fills the future up to the point where the truth becomes determined. If this is already the case before the actual evaluation time of the antecedent then this is where the fill-in stops.<sup>81</sup> The approach also can explain why the future uses of the present tense in the antecedent come without the certainty condition, that means why, in contrast to simple sentence in the present tense, future uses are acceptable even if (in the possibilities selected by the antecedent) it is not predetermined at the utterance time that the antecedent is true. While the epistemic reading selects for possibilities where the antecedent is true (and, hence, predetermination is necessary), the ontic reading makes the antecedent true. This is, of course, possible even without predetermination.

We have proposed in the previous section that also for the modals *MOLL* and *WOLL* an epistemic and an ontic reading can be distinguished. As for conditionals, the ambiguity is realized by the possibility of the modals to refer either to the epistemic or to the ontic interpretation function. This allows us to account for the variant of the puzzle of the shifted temporal perspective for modalities. The ontic reading predicts the observed shift of the reference time for tense in the scope of the modals. Additionally, we can account for the fact that modals can refer to the future without restriction to the certainty condition. The reason is again that because the ontic reading makes the formula in scope of the modal true it is not restricted to the presence of possibilities where the truth is predetermined at the utterance time. Notice that in this case future reference in the scope of a modal

---

<sup>81</sup>Notice, that at the same time we do not predict a backward shift for the temporal perspective if the antecedent refers to the past. The ontic update function never produces possibilities that are defined for a shorter period of time than the possibility to which the operation is applied.

is not made possible by a present tense (as in the antecedent of conditionals) but by the temporal properties of the modals itself. Because modals always refer to the revision-free variants of the update functions, we predict that if the truth of the formula in the scope of a modal is determined, the ontic reading comes down to the epistemic reading. Thus, in particular, we predict that for modal claims about the past and the present an ontic and an epistemic reading cannot be distinguished. This is in accordance with intuitions. We also predict that epistemic readings of modal statements about the future are rare, more precisely, they are only possible if the formula in the scope of the modal is taken to be possibly (in case of *MOLL*) or necessarily (in case of *WOLL*) predetermined. If this is the case, again the ontic reading and the epistemic reading are identical.

There is a lot of potential in the proposal made here that has to be explored in future work. For instance, one might think of accounting for other readings of the modalities in terms of the description of the ontic reading given here. One may go even further and think of using the ontic interpretation function in the description of the imperative mood which has also often been analyzed as prescribing how the world should be. To illustrate the potential that may lie in such an extension consider the case of conjunctive conditionals, exemplified in (156a) and (156b).

(156) a. Continue this behavior and I will fire you.

b. Tell him the truth and he will leave you.

Characteristic of these sentences is that *and* connects an imperative with a modal assertion. If you analyze imperatives as introducing in hypothetical contexts by applying the ontic interpretation function to the formula in scope of the imperative mood, then treating the sentence for the rest with the semantics proposed here will predict it to have a conditional reading. For illustration, consider the first of the two examples given above. An update with the sentence adds to the cognitive state a new hypothetical basic state where *hearer continues this behavior* has been made true in all possibilities where this can be made true. *And* says that this hypothetical context has to be updated with the sentence following the connective. The modal in this sentence performs a test on the hypothetical context. The test is successful if the hypothetical context supports (epistemic reading of the modal) or forces (ontic reading of the modal) the formula in scope of the modal.

Even though the approach presented here may have a lot of potential, it is also clear that a lot still has to be done. The proposal comes with a number of loose ends that need to be tied up in future work. Some of these loose ends were to be expected. The central goal of the present work was to formulate a formally precise approach to the semantics of English conditional sentences. The intention

was that the exact predictions made by such an approach will provide the basis for more elaborated empirical investigation. But it was not within the scope of the present work to undertake these empirical investigations. Indeed, at different places throughout the book we came across a number of very specific empirical questions about the meaning of certain conditional constructions that to our knowledge have not been addressed in the literature before. We hope that these questions will be answered in the future. Then, these answers can feed back into the theoretical work and help fine-tuning the present approach.

But the approach also raises a number of theoretical issues that have to be seen through in future work. Some of these open issues will be discussed below and possible answers sketched.

**Event semantics and the perfect.** One limitation of the approach mentioned already at the very beginning of this survey into the semantics of conditionals was the decision to not get involved in event semantics. There may lie a lot of potential, if not even a need, in extending the approach to event semantics and distinguish different aspect classes for verbs. For instance, there are good reasons to believe that event semantics would enable us to improve on the representation of causality in the present framework. The introduction of time made it necessary to make clear statements about the temporal relation between cause and effect to check the validity of causal laws. We have assumed here that the effect immediately follows the cause. Normal talk of cause and effect casts doubt on this assumption. Event semantics may allow for a more natural description of the relation between cause and effect.

There is another problem of the present approach where event semantics may help. As was said when we proposed a semantics for the perfect, the meaning we assumed for the perfect only captures its temporal properties and ignores the aspectual side of its meaning. This is a necessary consequence of our decision not to get involved with event semantics. It turns out that the proposed semantics for conditionals makes some unusual predictions for conditionals containing the perfect. Some of them, even though unexpected, may nevertheless be correct. But others really look like feedback of this limited approach to the meaning of the perfect.

One perhaps surprising prediction of the present approach is that it allows the perfect to talk about the past of some contextually given future time.<sup>82</sup> This seems to conflict with intuitions. Sentences like (157a) are – if at all – only marginally acceptable in English. But remember that if the perfect evaluates the formula in its scope at some point  $d$  in the future, then an epistemic update with this perfect formula will select those possibilities in the cognitive state in  $c_{\eta(c)}$  where it is already predetermined that the formula in the scope of the perfect is

---

<sup>82</sup>This is so because (i) the simple present can refer to the present as well as the future, and (ii) the present perfect is analyzed compositionally as *PERF* in the scope of *PRES*.

true. This suggests that we explain the unacceptability of (157a) as due to the fact that we find it difficult to think of finishing a dissertation as something that can now be determined to hold tomorrow. This is supported by the observation that the simple present pendant (157b) is unacceptable as well. But while for the simple present it is possible to find examples where this tense localizes the evaluation time in the future, it is difficult to find parallel examples for the present perfect. Consider, for instance, a sentence like (157c) where predetermination can be easily assumed. This sentence is still semantically anomalous for the native speakers we asked. More research on this question is certainly needed.

- (157) a. \*Tomorrow I have finished my dissertation.  
 b. \*Tomorrow I finish my dissertation.  
 c. ?Next year I have been working here for 30 years.

Matters become more complicated if we consider sentences where the present perfect occurs in the scope of an operator that makes reference to the ontic interpretation function. This is, for instance, the case in the antecedent of ontic conditionals. Also in this case the approach predicts that the perfect may refer to the past of some future reference time. But this time this is not a side-effect of the proposed semantics of the simple present, but of the way the function *AIntervene* is defined. As a consequence, it is not predicted that the formula in scope of the perfect has to be predetermined for the relevant past of the future. One consequence is that future readings of the present perfect in the antecedent of conditionals are claimed to be much more natural than in simple sentences like (157a) and not bound to predetermination. Indeed, the data support this prediction. Indicative conditionals with a present perfect in the antecedent that refers to the past of the future can easily be found.

- (158) a. If he still hasn't called next monday, we will contact the police.  
 b. If you have solved all these problems by next week, I will let you pass the examination.

Sometimes the application of *AIntervene* to sentences containing the perfect results in obvious mispredictions. Consider  $AIntervene_{M,g}^+(c, PERF_n(P)(d))$  where  $d$  and  $d_n$  refer to some point in the future with  $d_n < d$ . According to the interpretation rule of the perfect this update comes down to  $AIntervene_{M,g}^+(\bigcup_{t \in T} c[d_n/t][d_n < d], P(d_n))$ . Remember that for atomic formulas  $P(d_n)$ , *AIntervene* (normally) shifts the temporal perspective of the possibilities it selects forward to  $g_p(d_n)$ . Hence, an ontic update with the formula  $PERF_n(P)(d)$  where  $d$  and  $d_n$  lie in the future shifts the temporal perspective forward to  $g_p(d_n)$ . This prediction seems wrong. If the temporal perspective has to be shifted, then it should rather be

moved to the evaluation time of the perfect,  $g_p(d)$ , not to the evaluation time of the formula in its scope. The problem could easily be solved, if the update function *AIntervene* did not project through perfect formulas, but treated them as primitive. But why should this be the case? In approaches to the English perfect that take its aspectual properties into account, it has often been proposed that the perfect does not simply express that at some time in the past (of its evaluation time) the formula in its scope was true, but rather it asserts the existence of some result state or ‘Nachzustand’ at its evaluation time of an eventuality that fits the description in its scope. *AIntervene* applied to such a semantics for the perfect would make the existence of this result state true. As a consequence, there would also have to be the eventuality that produces this result state. But because it is the result state that is made true by *AIntervene*, the temporal perspective is shifted to the time of this result state – the evaluation time of the perfect. Such a more aspectual approach to the perfect, thus, would not make the problematic prediction described above.

Because we are already discussing the perfect, let us look at two other predictions made for the perfect that need empirical investigation. First, this approach predicts that there are also indicative conditionals with the past perfect. It turns out to be difficult to come up with natural examples of such conditionals. Sentence (159) might provide an example for this point.<sup>83</sup>

- (159) Peter woke up with a terrible hangover. He couldn’t remember what he had told the police. But one thing he was sure of: if he had told them his true name, then they would find him out.

Another prediction of the theory that needs to be investigated more closely is that in addition the perfect in subjunctive conditionals should allow for past-in-the-future readings. Because in case the simple past is interpreted as subjunctive, the semantic tense of the sentence is the simple present and this tense allows reference to future times, we predict that a subsequent perfect not interpreted as counterfactual mood should allow for an interpretation in the future as well. Below, we provide the translations of the examples (158a) and (158b) into the subjunctive. Again, the acceptability of such sentences needs to be investigated in the future.

- (160) a. If he still hadn’t called next monday, we would contact the police.  
 b. If you had solved all these problems by next week, I would let you pass the examination.

---

<sup>83</sup>This sentence is not an indicative conditional according to our initial definition of indicative conditionals given in Chapter 4, but it is indicative according to the approach to the English mood system proposed in this chapter.

**Update rules for modals and conditionals.** The approach presented here interprets conditionals and also the modals *WOLL* and *MOLL* as performing tests on cognitive states. They check whether the basic state of a cognitive state that has been introduced last fulfills certain conditions. If this is the case, then the cognitive state may be changed in that a new subordinate context is introduced. But no information about the actual world is gained by this update. If the test fails, the cognitive state is mapped to an absurd cognitive state. Thus, in a discourse conditional sentences and modal statements can convey information only in a very indirect manner. An unsuccessful update may lead to a conversation about why the speaker thinks that the test is successful. In this conversation information may be conveyed. But conditionals and modals do not do this by themselves. One may wonder whether this is correct, or whether conditionals and modal sentences can tell you something about the actual world directly. There are two difficulties in extending the provided test conditions to update conditions. The first difficulty is well-known and has also been discussed, for instance, in Veltman (2005). The interpretation of conditionals and also *WOLL* depends not only on the information available in the basic state to which the sentence is updated, but also on what are considered to be the laws. In this framework we treated the laws as fixed – something you cannot obtain new information about. Of course, this approach to laws may be extended in future work. We can learn new laws and we can give up on relations we previously thought to be laws. Conditionals and modal sentences may then not only provide information about the facts but also about what count as laws. It is a well-known problem of such an extension that the interpreter can then come into a situation where it is not clear whether he has to change what he believes to be the laws or what he believes to be the facts. Consider, for instance, that you believe some law saying that always if *A* is the case, then *B* will be the case as well. Assume, furthermore, that you are in a state, where you have no information about whether *A* or *B* holds. Now, somebody tells you *if B, then A*. There are in principle two distinct ways to update your information state. You may eliminate from the basic state to which the sentence is updated all possibilities where *A* is false and *B* is true, which would make the conditional a successful test with respect to your beliefs. But you may also conclude that a stronger law saying that *A* if and only if *B*, holds. Which update should you chose?

If you assume that the laws are fixed – as is done here – this problem does not occur and a unique update can be defined for *WOLL* and *IF*.<sup>84,85</sup> To simplify

---

<sup>84</sup>Respective update rules for *MOLL* do not make sense, because shrinking a basic state cannot turn an unsuccessful update with *MOLL*  $\psi$  into a successful one.

<sup>85</sup>We focus on the ontic readings of conditional and modal sentences here. The epistemic readings raise some issues of their own. Actually, it is even conceptually difficult to make sense of the idea of update for the epistemic reading. The epistemic reading works on the set of facts for which the interpreter has external evidence. Anything you can add is again a fact for which the interpreter is assumed to have external evidence. How can such information be provided by



matters a bit, let us forget for a moment about the introduction of hypothetical/subordinate contexts by *WOLL* and *IF* and take basic states  $R$  to be the objects update functions work on, in place of cognitive states. Then the test conditions provided above for these expressions can be simplified as follows.<sup>86,87</sup>

$$\begin{aligned} R[WOLL_n(\psi)(d)]_{M,g}^+ &= \begin{cases} R \text{ if } R \models \psi(d_n), \\ \emptyset \text{ if } R \not\models \psi(d_n). \end{cases} \\ R[IF_1 \psi, \phi]_{M,g}^+ &= \begin{cases} R \text{ if } AIntervene_{M,g}^+(R, \psi) \models \phi, \\ \emptyset \text{ if } AIntervene_{M,g}^+(R, \psi) \not\models \phi. \end{cases} \end{aligned}$$

Respective update conditions can then be formalized as follows.

$$\begin{aligned} R[WOLL_n(\psi)(d)]_{M,g}^+ &= \{p \in R \mid p \models \psi(d_n)\}, \\ R[IF_1 \psi, \phi]_{M,g}^+ &= \{p \in R \mid p \models \psi \Rightarrow p \models \psi \wedge \phi\}. \end{aligned}$$

As welcome as these update conditions may seem, there is still something to be desired. Assume that a conditional or a modal claim actually refers to some subordinate context. It seems only reasonable to assume that also in this case information is provided about the actual context, i.e.  $c_0$ . The description of update given above does not implement this idea. The update function only changes the subordinate context the conditional or the modal statement refers to. Intuitively, what we want is the following. Assume that the update eliminates some possibilities in some hypothetical state  $R$ . This hypothetical state was introduced by some update  $F$  to another basic state  $R'$ . Then we also should eliminate those possibilities in  $R'$  that by the update  $F$  resulted in the possibilities now eliminated in  $R$ . That means technically that somehow we have to be able to trace after update possibilities back to the (hypothetical) basic state they come from. How this can be made concrete has to be investigated in future work.

**Expectations.** Another open end left by the approach is the formalization of the expectations of an agent. We proposed that the ontic meaning of the modal *WOLL*, the ontic reading of conditionals, and the interpretation of the moods make reference to such expectations. Expectations were formalized as what can be derived from the facts and the general laws encoded in a law structure. We have already pointed out earlier that this formalization may turn out to be insufficient. Intuitively, it is clear that expectations can also rely on other sources of information not encoded in a law structure so far, like statistical laws, defaults,

---

conditionals? Consider again the example discussed above, where the interpreter assumes a law saying that always if  $A$ , then also  $B$ , but has no information about  $A$  and  $B$ . The interpreter is told *if B, then A*. Does this sentence provide external evidence for  $A \vee \neg B$ ?

<sup>86</sup>For the moment we also ignore the forward-extension of the evaluation time in the scope of a modal.

<sup>87</sup>The respectively simplified definitions for the update rules of simple sentences and the notions of subsistence and enforcement are straightforward.

graduated beliefs, etc. One topic for future research could be to extend the formalization of expectations used here to a description that comes closer to this intuitive notion. Then, one has to see whether such generalizations still lead to plausible predictions for the meaning of the moods, the modals, and conditionals proposed here.<sup>88</sup>

There is another sense in which the formalization of expectations given here may turn out to be inadequate. Our notion of expectations is a local concept, defined on the level of possibilities. The concept is lifted to the level of cognitive states  $c$  by quantifying over all possibilities in  $c_{\eta(c)}$ . As a consequence, there is no place for such a thing as epistemic expectations. More precisely it is not possible, that the truth value of some formula  $\psi$  is known to be defined ( $c \models \psi \vee \neg\psi$ ) without that the actual value being known ( $c \not\models \psi$ ,  $c \not\models \neg\psi$ ), but at the same time  $\psi$  is expected to be true ( $c \models \psi$ ). In other words, for facts about the past and the present, enforcement – and thereby the notion of expectation used here – reduces to support. This is at odds with our intuitive talk about expectations. It seems very natural that for some fact about the past we do not know whether it holds but nevertheless we expect it to hold. One may wonder whether perhaps the restricted information going into the calculation of expectations is responsible for this prediction. But at least part of the problem lies in the local definition of expectations. One may, therefore, think of introducing a concept of epistemic expectations based on the limited set of laws considered here:  $\psi$  is expected if it follows from facts believed in  $c_{\eta(c)}$  by the general laws encoded in a law structure. However, it is far from trivial to figure out how to make this idea precise. Is some fact expected if and only if it follows by laws from its immediate causes? What if the causes are not expected? One may also think of comparing the expectedness of possibilities by counting violations of causal laws:  $\psi$  is expected in a basic state  $R$  if it is locally expected in the possibilities in  $R$  with the smallest set of law-violations. But if then the expectedness of a certain fact is calculated based on this relation, do law-violations that do not concern the causal history of this fact also count? Do violations far back in history count less than recent violations? Independently of these problems, the introduction of an epistemic notion of expectedness may also make us lose some of the appealing predictions of the present approach. For instance, subjunctive sentences are now predicted to be unacceptable as update to the basic belief state  $c_0$ . The reason is that updates to  $c_0$  cannot violate the expectations of  $c_0$  (after update). This would no longer be the case if the present meaning of the subjunctive mood is tested with respect to an epistemic concept of expectations. Then it may be the case that some formula  $\psi$  is supported by  $c_0$  without being expected.

---

<sup>88</sup>In this connection we may as well call attention to some other point already discussed in Chapter 5. A law structure gives only a limited presentation of what agents may consider valid laws. An extension of what counts as law may be needed as well. Such a step may automatically lead to a more appropriate description of the expectations.

**Semantic function and cognitive processing.** If one closely considers the update conditions provided here for the modals and the connective *IF*, it attracts attention that the changes made to the cognitive state in case it passes the test condition are relevant for the test. The newly introduced hypothetical contexts have to be calculated to check the test condition. For instance,  $MOLL(\psi)$  introduces a hypothetical basic state containing the result of updating  $\psi$  to the basic state  $MOLL(\psi)$  applies to. At the same time, the modal tests whether this hypothetical update with  $\psi$  leads to the empty basic state.<sup>89</sup> This suggests that the introduction of the hypothetical state is only a side effect of calculating the test conditions. More generally, we want to propose that a difference has to be acknowledged between the semantic meaning associated with an expression and the way this meaning is processed. There are update effects that are not due to the semantic meaning, but due to its processing. One example for such interpretation effects is the introduction of hypothetical contexts.

To illustrate what is meant by this proposal let us make an excursion to programming. Assume that you want some primitive computer to calculate the function  $f(x) = x + x$ . The computer can read numbers off certain addresses of its memory, it can add 1 or subtract 1, and it can write numbers to addresses in its memory. Furthermore it can recognize when the value in some address is zero. The following pseudo-code describes a program this computer can process to double the value of some number written in address  $i$ .

```
begin;
j := i;
repeat;
  j := j - 1;
  i := i + 1;
until j = 0;
end;
```

We propose that the semantic value of some expression should be seen on a par with the function  $f(x) = x + x$ . This semantic value has to be distinguished from the cognitive mechanisms calculating the outcome of this functions for certain values. In the example, these mechanisms are the algorithm described by the program. The algorithm calculating the semantic value, in turn, relies on what the cognitive possibilities of the speaker/interpreter are, as does the program on what the computer hardware can and cannot do. The processing of the algorithm can have side effects that are totally independent of the function that the algorithm is implementing. For the example, executing this program does not only change the value in address  $i$  but also the value in address  $j$  that has been used for auxiliary calculations. In programming it is always very important to keep an eye on these

---

<sup>89</sup>Notice that the calculation of the relation  $c \models \psi$  involves the calculation of the update  $AIntervene_{M,g}^+(c, \psi)$ .

side effects of your algorithm. In semantics we have to do the same. We propose that the introduction of hypothetical contexts is an example of such a side effect of processing a semantic update. Hypothetical contexts are the outcome of an auxiliary calculation, stored in working memory. The working memory is what we called a cognitive state. To be more concrete, for the case of the modal *MOLL* the semantic value could be described as the function from basic states to basic states given below.

$$R[MOLL\psi]_{M,g} = \begin{cases} R & \text{if } R[\psi]_{M,g} \neq \emptyset, \\ \emptyset & \text{otherwise.} \end{cases}$$

Processing this function makes it necessary to write the update  $R[\psi]$  in a different address than the one where  $R$  is saved, because we may still need  $R$  to calculate the actual output basic state. A check of the value in this auxiliary address determines the output which is then written in the address of  $R$ . The semantic value has been calculated, but the calculation left the value of  $R[\psi]$  in some auxiliary address. These side effects of processing have then been facilitated by such other processes of language interpretation as, for instance, modal subordination.

Similar ideas to those developed here for the introduction of hypothetical contexts have been also formulated with respect to presupposition projection and anaphoric dependence. We propose that the introduction of hypothetical contexts can be explained by the update algorithm, as Stalnaker (1974), Karttunen (1974) and Heim (1992) propose that the projection behavior of presuppositions can be explained by the update algorithm: “... the phenomena of so-called presupposition projection are just by-products of how the CCP [context change potential, the author] of a complex sentence is composed from the CCPs of its parts.” (Heim, 1992: 185). Heim illustrates this point with the case of negation. It is well known that the negation of a sentence inherits the presuppositions of the sentence in the scope of the negation. The update rule for negation standardly assumed in dynamic semantics looks as follows.

$$R[\neg\psi]_{M,g} = R - R[\psi]_{M,g}$$

Presuppositions can then be interpreted as restricting the definedness of the update function  $[\psi]_{M,g}$  of a formula  $\psi$ . Given that the calculation of the update with  $\neg\psi$  involves the calculation of the update with  $\psi$ , it follows immediately that the restrictions on definedness of  $[\psi]_{M,g}$  project to the update function  $[\neg\psi]_{M,g}$ . But this outcome depends on the chosen description of the update rule. In principle, one could just as well define negation with a negative update function as is done in the proposal presented here. This would (for a two valued semantics) describe exactly the same semantic function. But this description would not involve the calculation of the update  $R[\psi]_{M,g}$ . Hence, the projection behavior of

the presupposition could not be explained.<sup>90</sup>

How can we make the distinction between semantic function and its processing transparent in the formalization? Let us sketch one possible approach and explain why we did not follow it here. First, one gives a description of the pure semantic content of the expression of the formal language. If the introduction of hypothetical contexts are not taken to be part of the semantic content, then, except for the moods, the type of the update function can be reduced to  $\langle\langle s, t \rangle, \langle s, t \rangle\rangle$ . Hence, the semantic meaning of a formula becomes a function from basic states to basic states. It is not difficult to formulate the new update rules. The only interesting cases are the modals and *IF*. For *MOLL* such a new update rules has been already provided above. *WOLL* and *IF* follow below.<sup>91</sup>

$$R[WOLL(\psi)(d)]_{M,g} = \begin{cases} R \text{ if } R \models \psi(d_n) \\ \emptyset \text{ if } R \not\models \psi(d_n) \\ \text{undefined if } \neg \forall p \in R : g_p(d_n) \geq g_p(d) \end{cases}$$

$$R[IF_1 \psi, \phi]_{M,g} = \begin{cases} R \text{ if } Intervene_{M,g}(R, \psi) \models \phi, \\ \emptyset \text{ if } Intervene_{M,g}(R, \psi) \not\models \phi \end{cases}$$

In addition to a description of the semantic meaning, we would also need a description of the cognitive implementation of the semantic meaning, i.e. of the algorithm for the calculation of its value. We should then be able to read processing effects, like the introduction of hypothetical context, off the algorithm implementing a certain semantic function.

One of the reasons why we did not follow this strategy here is that it is not at all trivial to describe the algorithms responsible for the calculation of the semantic values. Within the limitations of this dissertations there was no room to address this issue. A second reason is that this approach would bring back on stage another loose end of the approach. When looking at the ‘pure’ semantic meanings sketched above, it becomes quite clear that the semantics for the moods proposed in section 6.4 does not fit into this new picture. The mood cannot be interpreted as a basic state change function, because it makes reference to two different

---

<sup>90</sup>As an aside, applied to the given interpretation of negation, the claim we defend in this section would predict that this update rule for negation also introduces a hypothetical context with the content  $R[\psi]_{M,g}$ . There are some observations that support the assumption that also negation does introduce hypothetical contexts. For instance, negated sentences also allow for modal subordination (see example (161)). Indeed, some dynamic theories like DRT propose explicitly that the update with negation leads to the introduction of a hypothetical context with content  $R[\psi]_{M,g}$ .

(161) Peter didn’t drink any alcohol. He would have got sick.

<sup>91</sup>Only the ontic readings are considered here. There are no reasons for this choice, except brevity. The notion of enforcement has to be extended in a straightforward manner to apply to basic states.

basic states to calculate its output. Under this observation lurks a more general, conceptual problem. Given that the only thing the mood does is check whether the update with the sentence in their scope fulfills certain conditions, one may wonder whether they do actually operate on the semantic level, or rather, as has often be proposed, on the level of utterances, contributing to the assertion – or, more generally, speech act – conditions of utterances. However, because the mood does not appear at the top of the logical form of formulas but may be in the scope of sentential operators like *if* and *and*, one would need a compositional approach to the assertion conditions of utterances to implement this idea. Approaches along these lines have been made, for instance, by van der Sandt (1988, 1992), again in the context of theories of presupposition projection. Within this project we have not been able to work out such an approach within the framework introduced here. This is again an issue left for future work.

**Modal subordination** Related to the topic we ended with in the last paragraph, there is another loose end of the approach to the meaning of conditionals introduced in this chapter. As mentioned before, we have introduced only a very preliminary theory of inter-sentential modal subordination, that means of the mechanisms that determine the choice of basic state to which the semantic meaning of some utterance is updated. These mechanisms are claimed to lie entirely outside the semantic meaning of the utterances. We have proposed that the chosen context is by default the basic context  $c_0$ , where all information about the actual world is stored. If this update fails, then the last defined context  $c_{\eta(c)}$  is chosen. If also on this context the update is not successful, then the update in general fails. There are good reasons to doubt the general correctness of this preliminary proposal. Modal subordination may be partly determined by semantics. The semantic meaning of expressions like *then* and *in this case* may make it necessary to refer to subordinate contexts on a semantic level, possible by the introduction of variables for contexts. It may also be the case that the pragmatic mechanisms described here are not, or are only partly, correct. For instance, there is good evidence that sentences can also refer to contexts other than  $c_0$  and  $c_{\eta(c)}$ .

However, the theory developed here makes very precise predictions for intra-sentential modal subordination: within sentence boundaries the next formula is updated to the last index the cognitive state is defined for. This shows up in particular in the update conditions for the sentence connectives *and* and *if*. It is not very attractive to have a difference between inter- and intra-sentential subordination mechanism, but also not entirely unusual (see Veltman 1996). Much more serious, the intra-sentential mechanisms assumed here appear to lead to problems. As has been pointed out when the interpretation rule for *and* was introduced, this approach does not work for this connective. If the first conjunct introduces a new hypothetical context, this is normally not the context the second conjunct applies to. Consider, for instance, example (162).

(162) John may be in Stuttgart and he may be in Amsterdam.

The second conjunct certainly is not meant to be about those possibilities where John is in Stuttgart, as the present theory would predict. Actually, the sentence always leads to an absurd cognitive state in this theory. There are two ways to address this problem. First, one might propose that *and* has a different modal subordination behavior from *if* and should be treated on a par with modal subordination between independent sentences. A different option is to give up the proposal made for intra-sentential subordination. This is not so much a problem for the semantics of *if*. This interpretation rule can be restated without assuming the proposed theory of intra-sentential subordination. But it is a problem for the meaning of the moods. The description of the semantics of the moods makes very explicit reference to the index of the last introduced hypothetical state. This is particularly needed to explain the choice of mood of modal statements. While in general subjunctive sentences cannot refer to the basic context  $c_0$ , this is possible with sentences containing *might*. This is explained here by proposing that the subjunctive actually applies to the update with the formula in the scope of the modal – because this update defined the last introduced hypothetical context. The proposed theory of intra-sentential subordination provides an independent explanation for what makes this particular context special: it is the output context of an update. If we give up the proposal for intra-sentential subordination, then it becomes difficult to give a motivation for this behavior of the moods. It may in the end even become difficult to give a compositional approach to their meanings.<sup>92</sup> The mood needs to be able to easily address the context where the formula in the scope of the modal is updated. Maybe with more proper treatments of the introduction of hypothetical contexts and modal subordination this context is no longer available after the modal formula has been processed.

**The diachronic development of mood markings.** The approach presented in section 6.4 answers the central questions we asked at the beginning of the chapter. It provides a description of the semantic contribution of the syntactic tenses and the syntactic perfect and it explained the two observations concerning the temporal properties of conditionals that we have focussed on: the puzzle of the missing interpretation and the puzzle of the shifted temporal perspective. We may even be able to account for another observation mentioned when discussing the puzzle of the missing interpretation. It has often been observed that the temporal interpretation of the past tense is not only lost in conditionals but also in other linguistic contexts like hypothetical wishes, etc. We may propose that in all these contexts the simple past is interpreted as subjunctive mood. This hypothesis can only be verified by a close study of each of these contexts

---

<sup>92</sup>That means, if we maintain the description of the logical form of English sentences assumed here.

separately. However, when discussing the puzzling temporal properties of English conditionals and modals, we also touched on a number of other questions that we have not provided an answer for so far.

First, we have proposed here a synchronic theory of the meaning of mood, modality and conditionals in English. But the way English expresses conditionals, modals, and mood has changed remarkably over time. In section 6.2 we have discussed some theories that try to account for these changes. One very interesting objective of future research could be to see whether the synchronic proposal made here can be extended to a diachronic account of how and why English changed in its ways to express different types of conditionals etc.

Second, we have also noticed that some of the puzzling observations for the temporal behavior of English conditionals are not specific to this language. As James (1982) has shown, there is a large number of languages from different language families that share with English the puzzle of the missing interpretation for their past tense. Another interesting question to study in future work is whether the approach provided here can be extended with an explanation of this cross-linguistic pattern.

## 6.6 Summary

In this section we have introduced a compositional approach to the semantic meaning of English conditional sentences. This approach makes concrete proposals for the meaning of the tenses, the perfect and the modals *WOLL* and *MOLL* and for how these meanings contribute to the semantics of conditionals. The aim was to account for certain problematic observations concerning the temporal properties of English conditionals. These were (i) the puzzle of the missing interpretation, and (ii) the puzzle of the shifted temporal perspective.

The first puzzle refers to the observation that in subjunctive conditionals the simple past and the perfect markings in antecedent and consequent appear not to be interpreted. That is, the meaning of the conditionals does not show the temporal properties that would be expected given the standardly assumed meanings of the past tense and the perfect in English. Two ways to approach this problem can be distinguished in the literature. Firstly, it has been proposed that, even though it does not look that way, the tense and aspect morphology in subjunctive conditionals carries the same meaning as in simple sentences. Proposals along these lines follow roughly one and the same idea: in conditionals the past or the perfect do not shift the evaluation time of the antecedent and the consequent backward, but the evaluation time of the conditional as a whole. The price paid for being able to stick to the standard meaning for the tense- and aspect morphology in subjunctive conditionals is a logical form that does not follow the surface structure of the sentence. One problem for approaches



along these lines is that they can often only account for parts of the puzzle of the missing interpretation. That is, they can account for some of the past or perfect markings occurring in subjunctive conditionals, but not for all of them. Furthermore, we have argued that a description of the meaning of subjunctive conditionals as conditionals evaluated in the past is not able to correctly describe the truth conditions of such sentences. Alternatively, it has often been proposed that the simple past or the perfect have a mood/modality meaning in subjunctive conditionals. The criticism many proposals along this line have to bear is that they describe the meaning of the aspect and tense morphology in conditionals only in very vague terms. As a consequence, they make rather diffuse predictions for the semantics of these sentences and other constructions containing the same tense and aspect markings.

The approach developed here adopts the second approach to the puzzle of the missing interpretation. We have proposed that the simple past and the perfect obtain a mood interpretation in subjunctive conditionals. But in contrast to other approaches along this line, our proposal makes very specific claims about the meaning of the perfect and the simple past and the way they contribute their meanings to complex expressions. More particularly, we claim that English assertive sentences are obligatorily marked for mood. We distinguish three moods for assertions: an indicative mood, a subjunctive mood, and a counterfactual mood. We propose that the mood gives information about how the content of a sentence relates to the information about the actual world contained in the cognitive state to which the sentence is updated. In particular, the mood helps to determine when a statement gives information about a subordinate, hypothetical context. The indicative mood is described as demanding that the update with the sentence in its scope is consistent with the expectations, the subjunctive mood demands that the update is inconsistent with the expectations, and the counterfactual mood that the update is inconsistent with the information about the actual world contained in the cognitive state. We further proposed that the subjunctive and the counterfactual mood are marked in English using the simple past and the past perfect. Hence, according to the approach developed here, the form of the simple past and the perfect is ambiguous between a temporal/aspectual meaning and a mood meaning. This explains the puzzle of the missing interpretation.

The second puzzle that we wanted to account for, the puzzle of the shifted temporal perspective, concerns in the first instance the interpretation of the tenses in indicative conditionals. It has very often been proposed that the meaning of the tenses has a deictic element: they locate the evaluation time of the sentences they modify relative to the utterance time, the simple present locates the evaluation time at the utterance time (or in the non-past), the past locates this time before the utterance time. However, this appears to be falsified by indicative conditionals. For instance, a past tensed consequent in such conditionals can be evaluated

in the future of the utterance time. Something similar can be observed with respect to the interpretation of tense in subordinate relative clauses of modals. We have analyzed these observations as showing that in the consequent of conditionals and in modal contexts the reference time for the interpretation of tenses can be shifted forward. This shift is then explained as a natural consequence of the update conditions for the ontic reading of conditionals and modals.

One of the central claims we made was that a systematic distinction has to be made between an epistemic and an ontic reading of conditionals. This distinction applies to all types of conditionals, indicative conditionals as well as subjunctive conditionals or counterfactuals. The two readings are, however, not modeled by letting conditionals and modals refer to different modal bases, but by distinguishing two ways to update a sentence to an information state. The epistemic update takes a descriptive stance towards language use and interprets the sentence that has to be updated as providing information about the actual world. The ontic interpretation function assumes a prescriptive language use and makes the sentence that is to be updated true in the cognitive state to which the update applies. We are here only concerned with assertions. Thus, on the level of sentences always the epistemic update function is always applied. But we have proposed that there are lexical items whose epistemic interpretation can make reference to the ontic update function. Among these are the modals *WOLL* and *MOLL* and the sentence connective *IF*. One of the side effects of the working of the ontic interpretation function is that it can shift the temporal perspective of the possibilities it changes forward – and thereby the reference time for the interpretation of the tenses. This explains the puzzle of the shifted temporal perspective in conditional and modal contexts.

Besides the fact that the theory developed in this chapter can account for the puzzle of the missing interpretation and the puzzle of the shifted temporal perspective, it also can deal with the problems for the semantics of *would have* conditionals discussed in the previous chapter. The description of the epistemic and the ontic interpretation function proposed here is based on the proposal made for the respective readings of *would have* conditionals in Chapter 5. But the introduction of time into the model and the extended lexicon made it impossible to directly use the approach developed there. Some changes have been necessary. However, these changes have been conservative in that they do not affect the predictions made for the critical properties of *would have* conditionals discussed in Chapter 5.



## Appendix A

### Appendix to chapter 5

**A.0.1. FACT.** Let  $M = \langle W, F \rangle$  be a model for  $\mathcal{L}$ ,  $\langle \mathcal{B}, U \rangle$  a belief state of  $M$ , and  $\psi, \phi \in \mathcal{L}^0$ . The following claim does not hold for the function  $Learn$ .

$$\begin{aligned} & \text{If } Learn_M(\langle \mathcal{B}, U \rangle, \psi) \cap \llbracket \phi \rrbracket^M \neq \emptyset, \\ & \text{then } Learn_M(\langle \mathcal{B}, U \rangle, \psi \wedge \phi) \subseteq Learn_M(\langle \mathcal{B}, U \rangle, \psi). \end{aligned}$$

**Proof.** We proof the statement by providing a counterexample. The set of proposition letters from which the formal language  $\mathcal{L}$  is build up is  $\mathcal{P} = \{A, B, C\}$ . Take as belief state the tuple  $\langle \mathcal{B}, U \rangle$ , where  $U$  is the set of all interpretation functions for  $\mathcal{P}$  and  $\mathcal{B}$  the set  $\{\neg A, \neg B, \neg C\}$ . Figure A.1 describes the elements of  $U$  and the value of the function  $\mathcal{B}(w)$  for each of these elements.

$U$	$A$	$B$	$C$	$\mathcal{B}(w)$
$w_1$	0	0	0	$\{\neg A, \neg B, \neg C\}$
$w_2$	0	0	1	$\{\neg A, \neg B\}$
$w_3$	0	1	0	$\{\neg A, \neg C\}$
$w_4$	0	1	1	$\{\neg A\}$
$w_5$	1	0	0	$\{\neg B, \neg C\}$
$w_6$	1	0	1	$\{\neg B\}$
$w_7$	1	1	0	$\{\neg C\}$
$w_8$	1	1	1	$\emptyset$

Figure A.1: The elements of  $U$

Take  $\psi = A \vee B$  and  $\phi = \neg A \vee C$ . Then we obtain  $Learn_M(\langle \mathcal{B}, U \rangle, \psi) = Learn_M(\langle \mathcal{B}, U \rangle, A \vee B) = \{w_3, w_5\}$ . Hence,  $Learn_M(\langle \mathcal{B}, U \rangle, \psi) \cap \llbracket \phi \rrbracket^M = \{w_3\} \neq \emptyset$ . Thus, the assumption of the claim in fact A.0.1 is fulfilled. Furthermore, we calculate  $Learn_M(\langle \mathcal{B}, U \rangle, (\psi \wedge \phi)) = Learn_M(\langle \mathcal{B}, U \rangle, ((A \vee B) \wedge (\neg A \vee C))) = \{w_3, w_6\}$ .

But  $\{w_3, w_6\} \not\subseteq \{w_3, w_5\}$ . Hence,  $\text{Learn}_M(\langle \mathcal{B}, U \rangle, \psi \wedge \phi) \not\subseteq \text{Learn}_M(\langle \mathcal{B}, U \rangle, \psi)$ . This proves fact A.0.1.

**A.0.2. FACT.** Let  $M = \langle C, U \rangle$  be a model for  $\mathcal{L}$  and  $i \in I$  a partial interpretation of  $\mathcal{P}$ . The law closure  $\bar{i}$  of  $i$  is uniquely defined.

**Proof.** Assume that there are two partial interpretation functions  $i_1, i_2 \in I$  that fulfill the conditions (i) - (iii) of definition 5.6.13. We will show that in this case also  $i_3 \in I$  defined as  $i_1 \cap i_2$  fulfills the conditions (i)-(iii). From this it follows that if  $i_1 \neq i_2$ , then they cannot both be a law closure of  $i$ . This proves the claim.

Add (i) From  $i \subseteq i_1$  and  $i \subseteq i_2$  it follows  $i \subseteq i_3$ . Thus,  $i_3$  fulfills condition (i) of definition 5.6.13.

Add (ii) From  $i_3 \subseteq i_1$  it follows that  $\bigcap \{w \in U \mid i_3 \subseteq w\} \subseteq \bigcap \{w \in U \mid i_1 \subseteq w\}$ . The same holds for  $i_2$ . We conclude  $\bigcap \{w \in U \mid i_3 \subseteq w\} \subseteq i_1 \cap i_2 = i_3$ . On the other hand, we have  $i_3 \subseteq \{w \in U \mid i_3 \subseteq w\}$ . Thus,  $\bigcap \{w \in U \mid i_3 \subseteq w\} = i_3$ .  $i_3$  fulfills also condition (ii) of definition 5.6.13.

Add (iii) Assume that for  $P \in E$  with  $Z - P = \langle P_1, \dots, P_n \rangle$   $i(P)$  is undefined. Furthermore, assume that  $\forall k \in \{1, \dots, n\}$ ,  $i_3(P_k)$  is defined and  $f_P(i_3(P_1), \dots, i_3(P_n))$  is defined as well. From  $i_3 \subseteq i_1$  we can conclude that  $\forall k \in \{1, \dots, n\}$ ,  $i_1(P_k)$  is defined and  $i_1(P_k) = i_3(P_k)$ . This means that  $f_P(i_1(P_1), \dots, i_1(P_n))$  is defined and  $f_P(i_1(P_1), \dots, i_1(P_n)) = f_P(i_3(P_1), \dots, i_3(P_n))$ . Because  $i_1$  fulfills condition (iii) of definition 5.6.13, we also know  $f_P(i_1(P_1), \dots, i_1(P_n)) = f_P(i_3(P_1), \dots, i_3(P_n)) = i_1(P)$ . In the same way we reason that  $f_P(i_1(P_2), \dots, i_2(P_n)) = f_P(i_3(P_1), \dots, i_3(P_n)) = i_2(P)$ . From  $i_3 = i_1 \cap i_2$  it follows that  $i_3$  is defined for  $P$  and  $f_P(i_3(P_1), \dots, i_3(P_n)) = i_3(P)$ . Thus,  $i_3$  fulfills condition (iii) of definition 5.6.13.

**A.0.3. FACT.** Let  $M = \langle C, U \rangle$  be a model for  $\mathcal{L}$ ,  $w \in U$  a possible world, and  $\psi$  an element of  $\mathcal{L}^0$ . For all  $w, w' \in U$  it holds that if  $w =_1^w w'$ , then  $w = w'$ .

**Proof.** From  $w =_1^w w'$  it follows  $b_w \cap b_w = b_w = b_{w'} \cap b_w$ , i.e.  $b_{w'} \subseteq b_w$ . It additionally follows  $b_w - b_w = \emptyset = b_{w'} - b_w$ . But that means  $b_w = b_{w'}$ . Because of fact A.0.2 we know that, in consequence,  $\bar{b}_w = \bar{b}_{w'}$ . The definition of a basis allows us now to conclude  $w = w'$ .

## Appendix B

### Appendix to chapter 6

**B.0.4. LEMMA.** Let  $M = \langle S, \langle T, < \rangle, \langle C, U \rangle, I, \text{now} \rangle$  be a model for  $\mathcal{L}$  with  $U$  the set of all interpretation functions for  $\mathcal{P}$ ,  $g$  an assignment function, and  $p, p'$  possibilities for  $M$  and  $g$ .

- (i) If  $w_p \subseteq w_{p'}$ , then  $b_p \upharpoonright_{t_p} = b_{p'} \upharpoonright_{t_p}$ .
- (ii) For  $t \leq t_p, t_{p'}$ , if  $b_{p'} \upharpoonright_t = b_p \upharpoonright_t$ , then  $w_p \upharpoonright_t = w_{p'} \upharpoonright_t$ .
- (iii)  $\forall t \leq t_p \forall P \in B$ :  $b_p(P, t)$  is defined.

**B.0.5. FACT.** Let  $M$  be a model for  $\mathcal{L}$ ,  $g$  an assignment function,  $p$  a possibility for  $M$  and  $g$   $P \in \mathcal{P}$ , and  $d \in \text{VAR}_i$ .

$$\begin{aligned} AIntervene_{M,g}^+(p, P(d)) \neq \emptyset &\Rightarrow Intervene_{M,g}^+(p, P(d)) = AIntervene_{M,g}^+(p, P(d)) \\ AIntervene_{M,g}^-(p, P(d)) \neq \emptyset &\Rightarrow Intervene_{M,g}^-(p, P(d)) = AIntervene_{M,g}^-(p, P(d)) \end{aligned}$$

**Proof:** We show that the relation holds for positive update functions. The proof for the negative counterparts works along the same lines. In more details, we show that if  $AIntervene_{M,g}^+(p, P(d)) \neq \emptyset$ , then the following three subclaims hold.

- (1)  $\forall p' \in AIntervene_{M,g}^+(p, P(d)) \forall p'' \in \llbracket P(d) \rrbracket_{M,p}^+ : p'' \not\prec_1^p p'$ ,  
i.e.  $AIntervene_{M,g}^+(p, P(d)) \subseteq \text{Min}(\leq_1^p, \llbracket P(d) \rrbracket_{M,p}^+)$ .
- (2)  $\forall p'' \in \text{Min}(\leq_1^p, \llbracket P(d) \rrbracket_{M,p}^+) \exists p' \in AIntervene_{M,g}^+(p, P(d)) : p' \leq_2^p p''$ , i.e.  
 $AIntervene_{M,g}^+(p, P(d)) \supseteq \text{Min}(\leq_2^p, \text{Min}(\leq_1^p, \llbracket P(d) \rrbracket_{M,p}^+))$ .
- (3)  $\forall p', p'' \in AIntervene_{M,g}^+(p, P(d)) : p' \not\prec_2^p p''$ , i.e.  $AIntervene_{M,g}^+(p, P(d)) \subseteq \text{Min}(\leq_2^p, \text{Min}(\leq_1^p, \llbracket P(d) \rrbracket_{M,p}^+))$ .

**Add (1):** Assume that for  $p' \in AIntervene_{M,g}^+(p, P(d))$  and  $p'' \in \llbracket P(d) \rrbracket_{M,p}^+$  we have  $p'' <_1^p p'$ . By (i) it follows from  $w_p \subseteq w_{p'}$  that  $b_{p'} \cap b_p = b_p$ . From this we can conclude (using the assumption) that  $b_p \subseteq b_{p''}$ . This implies (using (ii))  $w_p \subseteq w_{p''}$ . On the other hand, we also conclude  $(b_{p''} - B) - b_p \subset (b_{p'} - B) - b_p$ . By assumption,  $p' \in \text{Min}(\preceq_2^p, \text{Min}(\preceq_1^p, \{q \in \llbracket P(d) \rrbracket_{M,g}^+ \mid w_p \subseteq w_q\}))$ . Thus, in particular,  $p' \in \text{Min}(\preceq_{M,g}^p, \{q \in \llbracket P(d) \rrbracket_{M,g}^+ \mid w_p \subseteq w_q\})$ . From this it follows  $\neg \exists p''' \in \{q \in \llbracket P(d) \rrbracket_{M,g}^+ \mid w_p \subseteq w_q\} : (b_{p'''} - B) - b_p \subset (b_{p'} - B) - b_p$ . But with  $p''$  we have found exactly such a  $p'''$ ! Contradiction.

**Add (2):** Assume  $p'' \in \text{Min}(\leq_1^p, \llbracket P(d) \rrbracket_{M,p}^+)$  and  $p'' \notin AIntervene_{M,g}^+(p, P(d))$ .

Case 1:  $w_p \subseteq w_{p''}$ . From  $p'' \in \text{Min}(\leq_1^p, \llbracket P(d) \rrbracket_{M,p}^+)$

we conclude together with  $w_p \subseteq w_{p''}$  that  $p'' \in \text{Min}(\preceq_1^p, \{q \in \llbracket P(d) \rrbracket_{M,p}^+ \mid w_p \subseteq w_q\})$ . In this case from  $p'' \notin AIntervene_{M,g}^+(p, P(d))$  it follows  $\exists p' \in AIntervene_{M,g}^+(p, P(d)) : p' \preceq_2^p p''$ .

Case 2:  $w_p \not\subseteq w_{p''}$ . From this assumption, together with  $p'' \in \text{Min}(\leq_1^p, \llbracket P(d) \rrbracket_{M,p}^+)$ , we can conclude  $(w_{p''} - b_{p''}) \cap (w_p - b_p) \subset w_p - b_p$ .<sup>1</sup> For any  $p' \in AIntervene_{M,g}^+(p, P(d))$  from  $w_p \subseteq w_{p'}$  it follows (by (iii)) that  $(w_{p'} - b_{p'}) \cap (w_p - b_p) = w_p - b_p$ . Hence,  $\forall p' \in AIntervene_{M,g}^+(p, P(d)) : p' <_2^p p''$ .

**Add (3):** From  $w_p \subseteq w_{p'}, w_{p''}$  we conclude (using (i))  $(w_{p''} - b_{p''}) \cap (w_p - b_p) = w_p - b_p = (w_{p'} - b_{p'}) \cap (w_p - b_p)$ . From  $p' \in AIntervene_{M,g}^+(p, P(d))$  we know  $\neg \exists p'' \in \{q \in \llbracket P(d) \rrbracket_{M,p}^+ \mid w_p \subseteq w_q\} : (w_{p''} - b_{p''}) - (w_p - b_p) \subset (w_{p'} - b_{p'}) - (w_p - b_p)$ . This proves the last subclaim.

---

<sup>1</sup>This step is not trivial. From  $w_p \not\subseteq w_{p''}$  alone this conclusion cannot be drawn. However, for those  $p''$  such that  $w_p \not\subseteq w_{p''}$  and  $(w_{p''} - b_{p''}) \cap (w_p - b_p) = w_p - b_p$  one can show that they are not minimal with respect to  $\leq_1^p$ . I leave this to the interested reader.

---

## Bibliography

- E.W. Adams. *The logic of conditionals. An application of probability to deductive logic*. Reidel, Dordrecht, 1975.
- E.W. Adams. Prior probabilities and counterfactual conditionals. In W.L. Harper and C.A. Hooker, editors, *Foundations of probability theory, statistical inference, and statistical theories of science*, volume 1, pages 1–22. Reidel, Dordrecht, 1976.
- C.E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50: 510–530, 1985.
- C.E. Alchourrón and D. Makinson. Hierarchies of regulations and their logic. In R. Hilpinen, editor, *New studies in deontic logic*, pages 125–148. Reidel, Dordrecht, 1981.
- C.E. Alchourrón and D. Makinson. On the logic of theory change: Contraction functions and their associated revision functions. *Theoria*, 48:14–37, 1982.
- M. Aloni. Free choice in modal contexts. In M. Weisgerber, editor, *Proceedings of the conference ‘SUB 7 -Sinn und Bedeutung’*, pages 25–37. University of Konstanz, 2002.
- M. Aloni. On choice-offering imperatives. In P. Dekker and R. van Rooy, editors, *Proceedings of the 14th Amsterdam Colloquium*, pages 57–62, Amsterdam, 2003.
- L. Alonso-Ovalle. Equal rights for every disjunct! Quantification over alternatives or pointwise context chance?, 2004. Handout, *Sinn und Bedeutung 9*, Nijmegen.
- A.R. Anderson. A note on subjunctive and counterfactual conditionals. *Analysis*, 12:35–38, 1951.



- A.C. Arregui. *On the accessibility of possible worlds: The role of tense and aspect*. PhD thesis, University of Massachusetts, Amherst, 2005.
- N. Asher and A. Lascarides. Bridging. *Journal of Semantics*, 15:83–113, 1998.
- N. Asher and E. McCready. Were, would, might, and a compositional account of counterfactuals. *Journal of Semantics*, 24(2):93–129, 2007.
- J. Atlas and S. Levinson. It-clefts, informativeness and logical form. In P. Cole, editor, *Radical Pragmatics*, pages 1–61. Academic Press, New York, 1981.
- A. Balke and J. Pearl. Probabilistic evaluation of counterfactual queries. In *Proceedings of the 12th national conference on artificial intelligence*, volume 11, pages 11–18, 1994.
- J. Bennett. Counterfactuals and temporal direction. *The Philosophical Review*, 93:7–89, 1984.
- J. Bennett. *A philosophical guide to conditionals*. Oxford University Press, Oxford, 2003.
- J. van Benthem. *The logic of time*. Reidel, Dordrecht, 1983.
- J. van Benthem. Semantic parallels in natural language and computation. In H. D. Ebbinghaus et al., editors, *Logic Colloquium '87*, pages 331–375, Amsterdam, 1989. Elsevier Science Publishers.
- P. Blackburn, M. de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.
- A. Bonomi and P. Casalegno. Only: association with focus in event semantics. *Natural Language Semantics*, 2:1–45, 1993.
- J.T. Boyland. A corpus study of *would* + *have* + Past-Participle in English. In R.H. Hogg and L. van Bergen, editors, *Historical Linguistics 1995*, volume 2, pages 1–17, Amsterdam/Philadelphia, 1998. Benjamins.
- B. Comrie. *Tense*. Cambridge University Press, Cambridge, 1985.
- B. Comrie. Conditionals: A typology. In E. Traugott et al., editors, *On conditionals*, pages 77–99. Cambridge University Press, Cambridge, 1986.
- C. Condoravdi. Temporal interpretation of modals. Modals for the present and for the past. In D. Beaver et al., editors, *The construction of meaning*, pages 59–87. CLSI Publications, 2002.
- C. Condoravdi. Moods and modalities for *will* and *would*. Handout, *The 14th Amsterdam Colloquium*, 2003.

- R. Crouch. *The temporal properties of English conditionals and modals*. PhD thesis, University of Cambridge, 1993.
- Ö. Dahl. The relation between past time reference and counterfactuality: A new look. In A Athanasiadou and R. Dirven, editors, *On conditionals again*, pages 97–112. John Benjamins, Amsterdam/Philadelphia, 1997.
- D. Denison. Counterfactual may have. In M. Gerritsen and D. Stein, editors, *Internal and external factors in syntactic change*, pages 229–256. Mouton, New York, 1992.
- P.B. Downing. Subjunctive conditionals, time order, and causality. In *Proceedings of the Aristotelian Society*, volume 59, pages 125–140, 1959.
- D.R. Dowty. *Word meaning and Montague grammar*. Reidel, Dordrecht, 1979.
- D.R. Dowty. Tenses, time adverbs and compositional semantic theory. *Linguistics and Philosophy*, 5:23–55, 1982.
- V.H. Dudman. Conditional interpretations of *if*-sentences. *Australian Journal of Linguistics*, 4:143–204, 1984.
- M. Enč. Towards a referential analysis of temporal expressions. *Linguistics and Philosophy*, 9:405–426, 1986.
- K. Fine. Review of Lewis (1973). *Mind*, 84:151–158, 1975.
- S. Fleischmann. Temporal distance: A basic linguistic metaphor. *Studies in Language*, 13:1–50, 1989.
- G. Forbes. Meaning-postulates, inference, and the relational/notional ambiguity. *Facta Philosophica*, 5:49–74, 2003.
- A. Frank. *Context dependence in modal constructions*. PhD thesis, University of Stuttgart, 1997.
- P. Gärdenfors. *Knowledge in Flux. Modeling the dynamics of epistemic states*. The MIT Press, 1988.
- G. Gazdar. *Pragmatics*. Academic Press, London, 1979.
- B. Geurts. Entertaining alternatives: Disjunctions as modals. *Natural Language Semantics*, 13(4):383–410, 2005.
- J. Ginzburg. Resolving questions, part 1. *Linguistics and Philosophy*, 18:459–527, 1995.

- F. Goodman. On the semantics of futurate sentences. In *Working papers in Linguistics*, volume 16, pages 76–89. Department of Linguistics, The Ohio State University, 1973.
- N. Goodman. *Fact, fiction and forecast*. The Bobbs-Merrill Company, Inc., Indianapolis/New York/Kansas City, 1955.
- G. Grahne. Updates and counterfactuals. In J.A. Allen, R. Fikes, and E. Sandewell, editors, *Principles of knowledge representation and reasoning: Proceedings of the 2nd international conference*, pages 269–276, San Mateo, California, 1991. Morgan Kaufmann.
- M. Green. Quantity, volubility, and some varieties of discourse. *Linguistics and Philosophy*, 18:83–112, 1995.
- H. P. Grice. Logic and conversation. In *Studies in the Way of Words*. Harvard University Press, Cambridge, 1989. Typescript from the William James Lectures, Harvard University, 1967.
- J. Groenendijk and M. Stokhof. *Studies in the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, University of Amsterdam, 1984.
- J. Groenendijk, M. Stokhof, and F. Veltman. Coreference and modality. In S. Lappin, editor, *Handbook of Contemporary Semantic Theory*, pages 179–216. Blackwell, Oxford, 1996.
- J.Y. Halpern and Y. Moses. Towards a theory of knowledge and ignorance. In *Proceedings 1984 Non-monotonic reasoning workshop*, pages 165–193, New Paltz, NY, 1984. American Association for Artificial Intelligence.
- C.L. Hamblin. Questions in Montague English. *Foundations of Language*, 10: 41–53, 1973.
- S.O. Hansson. New operators for theory change. *Theoria*, 55:114–132, 1989.
- R.M. Harnish. Logical form and implicature. In Bever T.G. et al., editors, *An integrated theory of linguistic ability*, pages 313–391. Crowell, New York, 1976.
- W.L. Harper. A sketch of some recent developments in the theory of conditionals. In W.L. Harper et al., editors, *IFS. Conditionals, belief, decision, chance, and time*, pages 3–38. Reidel, Dordrecht, 1981.
- I. Heim. Comments on Abusch's theory of tense. In H. Kamp, editor, *Ellipsis, tense and questions*, pages 141–170. University of Amsterdam, 1994.
- I. Heim and Kratzer A. *Semantics in generative grammar*. Blackwell, Cambridge, Mass., 1998.

- J. Hirschberg. *A theory of scalar implicature*. PhD thesis, University of Pennsylvania, 1985.
- W. van der Hoek et al. Persistence and minimality in epistemic logic. *Annals of Mathematics and Artificial Intelligence*, 27:25–47, 1999.
- W. van der Hoek et al. A general approach to multi-agent minimal knowledge. In M. Ojeda-Aciego, I.P. Guzman, G. Brewka, and L.M. Pereira, editors, *Proceedings JELLIA 2000*, LNAI 1919, pages 254–268, Heidelberg, 2000. Springer Verlag.
- L. Horn. *The semantics of logical operators in English*. PhD thesis, Yale University, 1972.
- L. Horn. Towards a new taxonomy of pragmatic inference: Q-based and R-based implicature. In D. Schiffrin, editor, *Meaning, Form, and Use in Context: Linguistic Applications, GURT84*, pages 11–42, Washington, 1984. Georgetown University Press.
- L. Horn. *A Natural History of Negation*. University of Chicago Press, Chicago, 1989.
- S. Iatridou. The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, 31(2):231–270, 2000.
- M. Ippolito. Presuppositions and implicatures in counterfactuals. *Natural Language Semantics*, 11:145–186, 2003.
- D. James. Past tense and the hypothetical: A cross-linguistic study. *Studies in Language*, 6:375–403, 1982.
- F. James. *Semantics of the English Subjunctive*. University of British Columbia Press, Vancouver, 1986.
- O. Jespersen. *The Philosophy of Grammar*. George Allen and Unwin, London, 1924.
- M. Joos. *The English verb*. University of Wisconsin Press, Madison, Wisc., 1964.
- N. Kadmon. The discourse representation of NPs with numeral determiners. In S. Berman et al., editors, *Proceedings of NELS 15*, pages 207–219, Amherst, 1985. GLSA, Linguistics Dept., University of Massachusetts.
- H. Kamp. Free choice permission. *Proceedings of the Aristotelian Society, N.S.*, 74:57–74, 1973.
- H. Kamp. The logic of historical necessity. unpublished manuscript, 1978.

- H. Kamp. Semantics versus pragmatics. In F. Guenther and S.J. Schmidt, editors, *Formal Semantics and Pragmatics of Natural Languages*, pages 255–287. Reidel, Dordrecht, 1979.
- H. Kamp and U. Reyle. *From Discourse to Logic*. Kluwer, Dordrecht, 1993.
- M. Kanazawa, S. Kaufmann, and S. Peters. On the lumping semantics of counterfactuals. *Journal of Semantics*, 22:129–151, 2005.
- L. Karttunen. Syntax and semantics of questions. *Linguistics and Philosophy*, 1: 3–44, 1977.
- L. Karttunen and S. Peters. Requiem for presupposition. In K. Whistler et al., editors, *Proceedings of the Third Annual Meeting of the Berkeley Linguistic Society*, pages 207–219, Berkeley, 1977. Berkeley Linguistics Society, University of California.
- H. Katsumono and A. Mendelzon. On the difference between updating a knowledge base and revising it. In J.F. Allen et al., editors, *KR'91: Principles of knowledge representation*, pages 387–394. Morgan Kaufmann, San Mateo, California, 1991.
- S. Kaufmann. Conditional truth and future reference. *Journal of Semantics*, 22 (3):231–280, 2005.
- A.M. Keller and M. Winslett Wilkins. On the use of an extended relational model to handle changing incomplete information. In *IEEE Trans. on Software Engineering*, volume 7 of *SE-11*, pages 620–633, 1985.
- K.B. Korb, C. Twardy, T. Handfield, and G. Oppy. Causal reasoning with causal models. Technical report, Monash University, 2005.
- A. Kratzer. Conditional necessity and possibility. In R. Bäuerle, U. Egli, and A. von Stechow, editors, *Semantics from different points of view*, pages 387–394. Springer, Berlin/Heidelberg/New York, 1979.
- A. Kratzer. The notional category of modality. In H.-J. Eikmeyer and H. Rieser, editors, *Words, worlds, and contexts*, pages 387–394. De Gruyter, Berlin/New York, 1981a.
- A. Kratzer. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, 10:201–216, 1981b.
- A. Kratzer. An investigation of the lumps of thought. *Linguistics and Philosophy*, 12:607–653, 1989.

- A. Kratzer. Conditionals. In A. von Stechow and D. Wunderlich, editors, *Semantics: An international handbook of contemporary research*, pages 651–656. De Gruyter, Berlin/New York, 1991a.
- A. Kratzer. Modality. In A. von Stechow and D. Wunderlich, editors, *Semantics: An international handbook of contemporary research*, pages 639–650. De Gruyter, Berlin/New York, 1991b.
- A. Kratzer. More structural analogies between pronouns and tenses. In Strolovitch D. and A. Lawson, editors, *Proceedings of SALT 8*, pages 620–633, Ithaca, NY, 1998. Cornell University, CLC Publications.
- J. van Kuppevelt. Inferring from topics. Scalar implicatures as topic-dependent inferences. *Linguistics and Philosophy*, 19:393–443, 1996.
- G. Lakoff. Presuppositions and relative wellformedness. In D. Steinberg and L. Jakobovits, editors, *Semantics*, pages 329–340. Cambridge University Press, Cambridge, 1971.
- R. Langacker. The form and meaning of the English auxiliary. *Language*, 54: 853–882, 1978.
- G. Leech. *Principles of Pragmatics*. Longman, London, 1983.
- O. Leirbukt. “Nächstes Jahr wäre er 200 Jahre alt geworden”: Über den Konjunktiv Plusquamperfekt in hypotetischen Bedingungsgefügen mit Zukunftsbezug. *Zeitschrift für Germanistische Linguistik*, 19:158–193, 1991.
- J. Lesli and S. Keeble. Do six-month-old infants perceive causality? *Cognition*, 25:265–288, 1987.
- D. Lewis. *Counterfactuals*. Blackwell, Oxford, 1973.
- D. Lewis. Counterfactual dependence and time’s arrow. *NOÛS*, 13:455–476, 1979a.
- D. Lewis. A problem about permission. In E. Saarinen et al., editors, *Essays in Honor of Jaakko Hintikka*, pages 163–175. Reidel, Dordrecht, 1979b.
- D. Lewis. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, 10:217–234, 1981a.
- D. Lewis. Probabilities of conditionals and conditional probability. In W.L. Harper et al., editors, *IFS. Conditionals, belief, decision, chance, and time*, pages 129–147. Reidel, Dordrecht, 1981b.

- S. Lindström and W. Rabinowicz. The Ramsey test revisited. In G. Grocco et al., editors, *Conditionals: From Philosophy to Computer Science*, pages 129–147. Clarendon Press, Oxford, 1995.
- J. Lyons. *Semantics*, volume 2. Cambridge University Press, Cambridge, 1977.
- Y. Matsumoto. The conversational condition on Horn scales. *Linguistics and Philosophy*, 18:21–60, 1995.
- J. McCarthy. Circumscription - a form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.
- J. McCarthy. Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence*, 28:89–116, 1986.
- J. McCawley. *Everything that linguists always wanted to know about Logic\**, volume 2. The University of Chicago Press, Chicago, 1993.
- A. Merin. Permission sentences stand in the way of Boolean and other lattice theoretic semantics. *Journal of Semantics*, 9:95–152, 1992.
- A. Merin. Information, relevance, and social decisionmaking. In L. Moss et al., editors, *Logic, Language, and Computation*, volume 2, pages 179–221. CSLI publications, Stanford, 1999.
- M. Morreau. Epistemic semantics for counterfactuals. *Journal of Philosophical Logic*, 21:33–62, 1992.
- J. Nerbonne. *German temporal semantics: Three dimensional tense logic*, volume 2. Garland Press, New York, 1985.
- F. R. Palmer. *Mood and modality*. Cambridge University Press, Cambridge, 1986.
- P. Parikh. A game-theoretical account of implicature. In Y. Vardi, editor, *Theoretical aspects of Rationality and Knowledge: TARK IV*, Monterey, California, 1992.
- B. Partee. Some structural analogies between tense and pronouns. *The Journal of Philosophy*, 70:601–609, 1973.
- J. Pearl. *Causality. Models, reasoning, and inference*. Cambridge University Press, Cambridge, 2000.
- D.M. Peterson and K.J. Riggs. Adaptive modelling and mindreading. *Mind & Language*, 14(1):80–112, 1999.
- R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik. *A communicative grammar of the English language*, volume 2. Longman, London/New York, 1985.

- F.P. Ramsey. General propositions and causality. In R.B. Braithwaite, editor, *Foundations of Mathematics and other logical essays by F.P. Ramsey*, pages 237–257. London, 1950.
- K.J. Riggs, D.M. Peterson, E.J. Robinson, and P. Mitchell. Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, 13:73–90, 1998.
- R. van Rooij. Permission to change. *Journal of Semantics*, 17:119–145, 2000.
- R. van Rooij. Questioning to resolve decision problems. *Linguistics and Philosophy*, 26:727–763, 2003.
- R. van Rooij. Utility, informativity and protocols. *Journal of Philosophical Logic*, 33:389–419, 2004.
- R. van Rooij and K. Schulz. Exhaustive interpretation of complex sentences. *Journal of Logic, Language, and Information*, 13:491–519, 2004.
- R. van Rooij and K. Schulz. *Only*: meaning and implicatures. In M. Aloni, A. Butler, and P. Dekker, editors, *The Semantics and Pragmatics of Questions and Answers*, pages 193–223. Elsevier, Amsterdam, 2007.
- M. Rooth. General propositions and causality. In S. Lappin, editor, *Focus*, pages 271–297. Blackwell Publishers, Malden, Mass., 1996.
- H. Rott. Moody conditionals: Hamburgers, switches, and the tragic death of an American president. In J. Gerbrandy et al., editors, *JFAK. Essays dedicated to Johan van Benthem on the occasion of his 50th birthday*, pages 98–112. Amsterdam University Press, Amsterdam, 1999.
- R. van der Sandt. *Context and presupposition*. Croom Helm, London, 1988.
- R. van der Sandt. Presupposition projection as anaphora resolution. *Journal of Semantics*, 9(4):333–377, 1992.
- U. Sauerland. Scalar implicatures of complex sentences. *Linguistics and Philosophy*, 27:367–391, 2004.
- K. Schulz. You may read it now or later: A case study on the paradox of free choice permission. Master’s thesis, University of Amsterdam, 2004.
- K. Schulz. A pragmatic solution for the paradox of free choice permission. *Synthese: Knowledge, Rationality and Action*, 147(2):343–377, 2005. see also Chapter 2.



- K. Schulz and R. van Rooij. Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy*, 29(2):205–250, 2006. see also chapter 3 of this book.
- H.A. Simon and N. Rescher. Cause and counterfactual. *Philosophy and Science*, 33:323–340, 1966.
- B. Skyrms. *Causal Necessity. A pragmatic investigation of the necessity of laws*. Yale University Press, New Haven, 1980.
- B. Skyrms. The prior propensity account of subjunctive conditionals. In W.L. Harper et al., editors, *IFS. Conditionals, belief, decision, chance, and time*, pages 259–265. Reidel, Dordrecht, 1981.
- B. Skyrms. *Pragmatics and empirism*. Yale University Press, New Haven, 1984.
- B. Skyrms. Adams conditionals. In E. Eells and B. Skyrms, editors, *Probability and conditionals. Belief revision and rational decision*, pages 13–29. Cambridge University Press, Cambridge, 1994.
- S. Soames. How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry*, 13:483–545, 1982.
- B. Spector. Scalar implicatures: Exhaustivity and Gricean reasoning? In B. ten Cate, editor, *Proceedings of the ESSLLI 2003 Student session*, Vienna, 2003.
- R. Stalnaker. A theory of conditionals. In J.W. Cornman et al., editors, *Studies in Logical Theory: essays*, pages 98–112. Blackwell, Oxford, 1968.
- R. Stalnaker. Assertion. In *Syntax and Semantics*, volume 9, pages 78–95. Academic Press, New York, 1978.
- R. Stalnaker. Letter to David Lewis. In W.L. Harper et al., editors, *IFS. Conditionals, belief, decision, chance, and time*, pages 151–152. Reidel, Dordrecht, 1981a.
- R. Stalnaker. Probability and conditionals. In W.L. Harper et al., editors, *IFS. Conditionals, belief, decision, chance, and time*, pages 107–128. Reidel, Dordrecht, 1981b.
- A. von Stechow. Semantisches und morphologisches Tempus: Zur temporalen Orientierung von Einstellungen und Modalen. *Neue Beiträge zur Germanistischen Linguistik*, 4(2), 2005.
- A. von Stechow and T.E. Zimmermann. Term answers and contextual change. *Linguistics*, 22:3–40, 1984.

- S. Steele. Past and irrealis: Just what does it all mean? *International Journal of American Linguistics*, 41:200–217, 1975.
- T. Stowell. Tense and modals. to appear.
- A. Szabolcsi. The semantics of topic-focus articulation. In J. Groenendijk et al., editors, *Formal methods in the study of language*, pages 513–540. Mathematisch Centrum, Amsterdam, 1981.
- P. Tedeschi. Some evidence for a branching-futures semantic model. In P. Tedeschi and A. Zaenen, editors, *Syntax and Semantics: Tense and Aspect*, volume 14. Academic Press, New York, 1981.
- R.H. Thomason. Combinations of tense and modality. In D. Gabbay and F. Günthner, editors, *Handbook of Philosophical logic: Extensions of Classical Logic*, pages 135–165. Reidel, Dordrecht, 1984.
- R.H. Thomason and A. Gupta. A theory of conditionals in the context of branching time. In W.L. Harper et al., editors, *IFS. Conditionals, belief, decision, chance, and time*. Reidel, Dordrecht, 1981.
- P. Tichy. A counterexample to the Stalnaker-Lewis analysis of counterfactuals. *Philosophical Studies*, 29:271–273, 1976.
- F. Veltman. Prejudices, presuppositions and the theory of counterfactuals. In J. Groenendijk et al., editors, *Amsterdam Papers in Formal Grammar*, volume 1. Centrale Interfaculteit, Universiteit van Amsterdam, 1976.
- F. Veltman. *Logics for conditionals*. PhD thesis, University of Amsterdam, 1985.
- F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3): 221–261, 1996.
- F. Veltman. Making counterfactual assumption. *Journal of Semantics*, 22:159–180, 2005.
- F.Th. Visser. *An historical syntax of the English language*. Brill, Leiden, 1973.
- J. Wainer. *Uses of nonmonotonic logic in natural language understanding: generalized implicatures*. PhD thesis, Pennsylvania State University, 1991.
- G.H. von Wright. *An Essay on Deontic Logic and the Theory of Action*. North-Holland Publishing Company, Amsterdam, 1968.
- H. Zeevat. Questions and exhaustivity in update semantics. In H. Bunt et al., editors, *Proceedings of the International Workshop on Computational Semantics*, Tilburg, 1994. Institute for Language Technology and Artificial Intelligence.

- T.E. Zimmermann. Free choice disjunction and epistemic possibility. *Natural Language Semantics*, 8:255–290, 2000.

---

# Index

- $UV(p)$ , 60
- $W_{M,g}$ , 209, 217
- $\div$ , 231
- $\nabla p$ , 14
- $\bar{i}$ , 144
- $\bar{w}$ , 221
- $\triangle p$ , 14
- $dom(s)$ , 208
- $g_s$ , 209
- $p \rightsquigarrow p'$ , 225
- $p \rightarrow p'$ , 226
- $t_s$ , 208
- $w_s$ , 208
- $\mathcal{S}$ , 15
  
- antecedent, 77
- assignment, 209
- assignment function, 209
  
- background variable, 104, 141
- backtracking, 87
- basic state, 210
- basis
  - epistemic reading
    - chapter 5, 132
    - chapter 6, 248
  - ontic reading
    - chapter 5, 144
    - chapter 6, 222
  - Veltman, 98
  
- belief revision
  - chapter 5, 135
  - chapter 6, 250
- belief state, 132
  
- causal model
  - formal, 104
  - informal, 103
- causal structure
  - formal, 141
  - informal, 141
- cognitive state, 210
  - extended, 248
- comparing competence, 67
- comparing relevant knowledge, 64
- competence, 22
- conditional sentence, 77
- consequent, 77
  
- determinism
  - determinism of laws, 216
  - Pearl's determinism, 106, 115
- dishonest sentences, 20
- domain of interpretation, 212
  
- endogenous variable, 104, 141
- enforcement, 227
- epistemic reading of conditionals
  - chapter 5
    - formal, 132
    - informal, 127

- chapter 6, 247
- epistemic update function
  - ALearn*, 214
  - Learn*
    - chapter 5, 135
    - chapter 6, 250
- evaluation time, 80
- exhaustive interpretation, 37
  - basic case, 49
  - by Groenendijk & Stokhof, 44
  - dynamic variant, 52
  - with relevance, 59
- expression of  $\mathcal{L}$ , 203
- faithful, 249
- follow
  - chapter 5, 141
  - chapter 6, 225
- formal language
  - chapter 2
    - $\mathcal{L}$ , 14
    - $\mathcal{L}^0$ , 14
  - chapter 5
    - $\mathcal{L}$ , 104
    - $\mathcal{L}^0$ , 131, 140
    - $\mathcal{L}^>$ , 131
    - $\mathcal{L}^{\gg}$ , 140
  - chapter 6
    - $\mathcal{L}$ , 201
- formula of  $\mathcal{L}$ , 203
- frame, 14
- free choice inferences, 11, 15
  - (D1), 15
  - (D2), 15
  - (D3), 15
  - (D4), 15
  - (D5), 15
- Gricean principle, 17, 62
- Gricean principle plus competence, 67
- index of a cognitive state, 202
- indicative conditional, 79
- information state, 52
- interpretation functions
  - chapter 2
    - grice*, 17
  - chapter 3
    - $[\cdot]$ , 44
    - $\cdot[\cdot]$ , 51
    - CIRC, 48
    - eps*, 67
    - exh<sub>GS</sub>*, 44
    - exh<sub>dyn</sub>*, 52
    - exh<sub>rel</sub>*, 59
    - exh<sub>std</sub>*, 49
    - grice*, 63
  - chapter 5
    - Intervene*, 145
    - Learn*, 135
    - $[[\cdot]]$ , 107, 131, 141
  - chapter 6
    - AIIntervene*, 217, 223
    - ALearn*, 214
    - Intervene*, 251
    - Learn*, 250
- interpretation rules for conditionals
  - backshift interpretation rule, 167
  - Pearl, 107
  - premise semantics, 96
  - Ramsey, 121
  - Stalnaker, 122
  - Tedeschi, 165
  - The Gibbart & Harper Causal Paradigm, 125
  - The Gibbart & Harper similarity approach, 125
  - Veltman, 99
- interpretation rules for operators
  - COUNT*, 240
  - IND*, 236
  - MOLL*
    - epistemic reading, 232
    - ontic reading, 231
  - PAST*, 228
  - PERF*, 230
  - PRES*, 228

- SUBJ*, 236
- IF*
  - epistemic reading, 247
  - ontic reading, 247
- WOLL*
  - epistemic reading, 232
  - ontic reading, 231
- ZERO*, 228
  - basic logical operators, 227
- Intervention
  - Pearl, 107
- law closure
  - chapter 5, 144
  - chapter 6, 221
- law structure, 208
- lexicon, 206
  - modified, 246
- minimal models, 4, 48
- minimality operator, 4, 135, 217
- model
  - chapter 2, 14
  - chapter 5
    - epistemic reading, 131
    - ontic reading, 141
  - chapter 6, 207
- notions of entailment
  - chapter 2
    - $\models$ , 15, 17
    - $\models^+$ , 24
    - $\models^0$ , 18
    - $\models^g$ , 32
    - $\models^n$ , 21
    - $\models^{g+}$ , 32
    - $\models$ , 14
- notions of truth
  - chapter 2
    - $s \models \phi$ , 14
  - chapter 5
    - $M, w \models \psi$ , 131, 141
  - chapter 6
    - $c \models \psi$ , 226
    - $c \models \psi$ , 227
- ontic reading of conditionals
  - chapter 5
    - formal, 143
    - informal, 127
  - chapter 6, 247
- ontic update function
  - AI*ntervene
    - final version, 223, 225
    - preliminary version, 217
  - Intervene*
    - chapter 5, 145
    - chapter 6, 251, 253
- orders
  - chapter 2
    - $\preceq^+$ , positive information order, 24
    - $\preceq^0$ , basic order, 18
    - $\preceq^g$ , general information order, 32
    - $\preceq^n$ , objective information order, 21
  - chapter 3
    - $<_P$ , 48, 52
    - $<_R$ , 59
    - $\cong_{P,A}$ , 64
    - $\equiv_{P,A}^*$ , 64
    - $\leq_{P,A}^*$ , 64
    - $\preceq_{P,A}$ , 64
    - $\sqsubseteq_{P,A}$ , 67
  - chapter 5
    - $\leq_1^w$ , 145
    - $\leq_2^w$ , 145
    - $\leq^{(B,U)}$ , 134
  - chapter 6
    - $\leq_1^p$ , 251
    - $\leq_2^p$ , 251
    - $\preceq_1^p$ , 223
    - $\preceq_2^p$ , 223
- paradox of free choice permission, 7, 8

- possibility
  - chapter 3, 52
  - chapter 6
    - preliminary definition, 209
    - with predetermination, 217
- possible world, 131, 141
- predicate circumscription, 47, 48
- quantification problem, 47
- readings of answers
  - domain restriction reading, 41
  - fine-grainedness dependence, 42
  - mention all reading, 41
  - mention some reading, 41
  - scalar reading, 41
- reference time, 80
- revision
  - global revision for belief states, 122
  - local revision for belief states, 123
  - local revision for worlds, 123
- rootedness
  - ontic reading, 141
  - Pearl, 104
- satisfaction
  - chapter 2, 14
  - chapter 5, 132
- sentence of  $\mathcal{L}$ , 205
- sentence schemes
  - [4], 14
  - [5], 14
  - $[C_1]$ , 22
  - $[C_2]$ , 22
  - $[D]$ , 14
- state, 14
- subjunctive conditional, 77
- subsistence, 226
- support, 226
- types, 202
- universe
  - epistemic reading, 132
  - ontic reading, 141
- utility value, 60
- utterance time, 80
- vocabulary, 202
- would conditional, 78
- would have conditional, 78

---

## Samenvatting

In de studie van de betekenis van taal onderscheiden wij de letterlijke betekenis – de betekenis die een uitdrukking heeft onafhankelijk van zijn gebruik – en de betekenis die een uitdrukking kan krijgen door de situatie waarin de uitdrukking is gebruikt. Bijvoorbeeld, de zin *Dit is mijn echtgenoot* betekent letterlijk dat de aangewezen persoon in een bepaalde wettelijke relatie staat met de spreker. Maar als je deze zin in een bar gebruikt tegenover een man die behoorlijk lastig begint te worden, dan kan de zin in deze situatie ook betekenen dat je met rust gelaten wilt worden. De letterlijke betekenis van een uitdrukking noemen wij ook zijn *semantische* betekenis. De betekenis die een uitdrukking kan krijgen door de situatie waarin de uitdrukking wordt geuit heet zijn *pragmatische* betekenis.

Een centrale vraag in de studie van de betekenis van taal is waar precies de grenslijn tussen semantiek en pragmatiek moet worden getrokken. Voor veel concrete aspecten van de betekenis van uitdrukkingen is het nog niet duidelijk of we ze als deel van de letterlijke betekenis van de uitdrukking moeten begrijpen, of als een effect van de interactie met de utings-contexts. In dit proefschrift worden drie bekende fenomenen bestudeerd waarvoor deze vraag nog open is. Voor alle drie de fenomenen wordt een concreet antwoord op de vraag voorgesteld in de vorm van een formeel uitgewerkte theorie, die het fenomeen als semantische of pragmatische inferentie verklaart.

Het eerste fenomeen dat we in dit proefschrift bestuderen is dat van de *vrije keuze inferenties* die vaak in verband met disjunctieve modale zinnen optreden. Bijvoorbeeld, een zin als (163) kan zo worden geïnterpreteerd dat de geadresseerde zowel een appel als ook een peer mag pakken. Hij heeft dus een vrije keuze tussen twee opties.

(163) Je mag een appel of een peer pakken.

Gangbare theoriën ter beschrijving van de semantische betekenis van (163) kunnen niet verklaren waar deze vrije keuze vandaan komt. In het tweede hoofdstuk van het voorliggende proefschrift bouwen wij voort op het idee, dat vrije



keuze inferenties deel van de pragmatische betekenis van zinnen als (163) zijn. We analyseren deze inferenties als conversationele implicaturen in de betekenis van Grice (1989). Een van de centrale zwaktes van Grice's theorie over de conversationele implicatuur is dat als gevolg van zijn algemene karakter de theorie geen concrete voorspellingen kan maken. Daarom ontwikkelen we hier eerst een gedeeltelijke formalisering van Grice's theorie en laten dan zien dat de formalisering vrije keuze inferenties correct kan voorspellen.

Het tweede fenomeen dat in dit proefschrift wordt besproken is de speciale manier waarop we normaliter antwoorden op vragen interpreteren. Ter illustratie, Bart's antwoord in voorbeeld (164) wordt vaak begrepen als een volledig antwoord op de vraag van Anna: niet alleen als de bewering dat Jan en Marie op het feestje komen, maar bovendien dat zij de enige twee personen zijn die komen.

(164) Anna: Wie komen er op het feestje?

Bart: Jan en Marie.

Deze lezing van antwoorden wordt hun uitputtende of *exhaustieve* interpretatie genoemd. In het derde hoofdstuk van het proefschrift wordt een formele beschrijving van dit fenomeen voorgesteld dat voor vele bekende vragen over de exhaustieve interpretatie een antwoord biedt. We stellen voor om ook de exhaustieve interpretatie van antwoorden als pragmatisch fenomeen, in het bijzonder als conversationele implicatuur te analyseren. We laten zien dat de formalisering van Grice's theorie voor conversationele implicaturen voorgesteld in hoofdstuk twee ook de exhaustieve interpretatie van antwoorden als implicatuur voorspeld.

In de laatste drie hoofdstukken van het proefschrift wordt de betekenis van conditionele zinnen in het Engels besproken. In het bijzonder zoeken wij in dit gedeelte van het proefschrift een verklaring voor de schijnbare discrepantie tussen de vorm van Engelse conditionele zinnen en hun temporele eigenschappen. Bijvoorbeeld, in subjunctieve conditionele zinnen zoals (165) draagt het finite werkwoord in de eerste deelzin (de *antecedent*) een markering voor de verleden tijd (*simple past*). Maar deze deelzin kan niet worden geïnterpreteerd als verwijzend naar het verleden.

(165) If you asked him, Peter would help you.

In het proefschrift ontwikkelen wij een benadering van de semantische betekenis van Engelse conditionele zinnen die hun temporele eigenschappen compositioneel van de betekenis van hun delen afleidt. Dus, in tegenstelling tot de eerste twee onderwerpen van de proefschrift, is het in dit geval de semantiek die verantwoordelijk wordt geacht voor het fenomeen dat we willen verklaren.

Maar voordat we beginnen met het ontwikkelen van een compositionele semantiek voor tempus markeringen in conditionele zinnen, wordt in hoofdstuk vijf de logische relatie tussen antecedent (de door *if* ingeleide bijzin) en consequent (de hoofdzin) van conditionele zinnen onder de loep genomen. De reden is dat eerst enkele vragen over de interpretatie van in het bijzonder counterfactische conditionele zinnen moeten worden beantwoord, voordat we aan een analyse van de temporele eigenschappen van deze zinnen kunnen beginnen. We ontwikkelen in hoofdstuk vijf een tijd-vrije semantiek voor formele zinnen van de vorm  $A > C$ , waarbij  $A$  en  $C$  voor de antecedent en de consequent van een counterfactische conditionele zin staan en  $>$  de conditionele connectief symboliseert. Daarna wordt in hoofdstuk zes het tijd-vrije raamwerk uit hoofdstuk vijf uitgebreid met (i) een gedetailleerdere structurele analyse van conditionele zinnen die modale en temporele markeringen onderscheidt, en (ii) een compositionele theorie voor de interpretatie van de complexe logische vorm van conditionele zinnen. We laten zien dat deze uitbreiding een verklaring voor de in dit proefschrift bestudeerde temporele eigenschappen van conditionele zinnen oplevert.

Naast hun relevantie voor de discussie over de scheidslijn tussen semantiek en pragmatiek is er nog een ander aspect dat alle drie onderwerpen van het proefschrift delen. In alle drie gevallen wordt de interpretatie van zinnen beschreven met gebruik van *minimale modellen*. Ter verduidelijking, laten we aannemen dat we een functie  $I$  hebben die aan zinnen  $\psi$  van een formele taal  $\mathcal{L}$  interpretaties toewijst. Meer in detail associeert de functie  $I$  elementen van  $\mathcal{L}$  met deelverzamelingen van een klasse  $M$  van modellen voor  $\mathcal{L}$ -zinnen. Dan kunnen we een sterkere interpretatie functie  $I^*$  beschrijven, die zinnen  $\psi \in \mathcal{L}$  op een subset van  $I(\psi)$  afbeeldt. Gegeven een ordening  $\leq$  op  $M$  kunnen we deze deelverzameling bijvoorbeeld bepalen als de verzameling van  $\leq$ -minimale modellen in  $I(\psi)$  :  $I^*(\psi) = \text{Min}(\leq, I(\psi))$ .

Zulk versterkingen van een basis interpretatiefunctie staan centraal in de formele benaderingen van alle drie de fenomenen die in dit proefschrift worden bestudeerd: vrije keuze inferenties, exhaustieve interpretatie en de betekenis van conditionele zinnen in het Engels. In het tweede en het derde hoofdstuk worden minimale modellen voor het beschrijven van de pragmatische betekenis van zinnen gebruikt. Ze spelen een centrale rol in de formalisering van Grice's (1989) theorie van conversationele implicaturen die we in hoofdstuk twee voorstellen. In dit verband beschrijft de functie  $I$  de semantische betekenis van zinnen en is  $I^*$  een versterking van de semantische betekenis met pragmatische informatie. In het tweede gedeelte van het proefschrift over temporele eigenschappen van conditionele zinnen worden minimale modellen al voor de formele beschrijving van de semantische betekenis van zinnen gebruikt. Zoals gebruikelijk in de literatuur nemen wij aan dat een conditionele zin met antecedent  $A$  en consequent  $C$  waar is in een mogelijke wereld  $w$ , als in alle mogelijke werelden waar het antecedent waar is en die het meest op  $w$  lijken ook het consequent waar is. Vergelijkbaarheid van

mogelijke werelden wordt dan beschreven met behulp van een ordening  $\leq$  tussen werelden: we zeggen dat wereld  $w_1$  kleiner is dan wereld  $w_2$  met betrekking tot wereld  $w$  als  $w_1$  meer op  $w$  lijkt dan  $w_2$ . Ook in deze samenhang is  $I$  (een abstracte versie van) een semantische interpretatie functie. Maar ook  $I^*$  is een semantische interpretatie functie: we beschrijven de operatie  $*$  als deel van de semantische betekenis van de conditionele connectief. Een centrale bijdrage van het werk gepresenteerd in dit proefschrift ligt in de manier waarop de vergelijkbaarheid van mogelijke werelden – en dus de operatie  $*$  – wordt beschreven. We stellen dat wetten, in het bijzonder causale wetten, hierbij een centrale rol spelen.

*Titles in the ILLC Dissertation Series:*

ILLC DS-2001-01: **Maria Aloni**

*Quantification under Conceptual Covers*

ILLC DS-2001-02: **Alexander van den Bosch**

*Rationality in Discovery - a study of Logic, Cognition, Computation and Neuropharmacology*

ILLC DS-2001-03: **Erik de Haas**

*Logics For OO Information Systems: a Semantic Study of Object Orientation from a Categorical Substructural Perspective*

ILLC DS-2001-04: **Rosalie Iemhoff**

*Provability Logic and Admissible Rules*

ILLC DS-2001-05: **Eva Hoogland**

*Definability and Interpolation: Model-theoretic investigations*

ILLC DS-2001-06: **Ronald de Wolf**

*Quantum Computing and Communication Complexity*

ILLC DS-2001-07: **Katsumi Sasaki**

*Logics and Provability*

ILLC DS-2001-08: **Allard Tamminga**

*Belief Dynamics. (Epistemo)logical Investigations*

ILLC DS-2001-09: **Gwen Kerdiles**

*Saying It with Pictures: a Logical Landscape of Conceptual Graphs*

ILLC DS-2001-10: **Marc Pauly**

*Logic for Social Software*

ILLC DS-2002-01: **Nikos Massios**

*Decision-Theoretic Robotic Surveillance*

ILLC DS-2002-02: **Marco Aiello**

*Spatial Reasoning: Theory and Practice*

ILLC DS-2002-03: **Yuri Engelhardt**

*The Language of Graphics*

ILLC DS-2002-04: **Willem Klaas van Dam**

*On Quantum Computation Theory*

ILLC DS-2002-05: **Rosella Gennari**

*Mapping Inferences: Constraint Propagation and Diamond Satisfaction*

- ILLC DS-2002-06: **Ivar Vermeulen**  
*A Logical Approach to Competition in Industries*
- ILLC DS-2003-01: **Barteld Kooi**  
*Knowledge, chance, and change*
- ILLC DS-2003-02: **Elisabeth Catherine Brouwer**  
*Imagining Metaphors: Cognitive Representation in Interpretation and Understanding*
- ILLC DS-2003-03: **Juan Heguiabehere**  
*Building Logic Toolboxes*
- ILLC DS-2003-04: **Christof Monz**  
*From Document Retrieval to Question Answering*
- ILLC DS-2004-01: **Hein Philipp Röhrig**  
*Quantum Query Complexity and Distributed Computing*
- ILLC DS-2004-02: **Sebastian Brand**  
*Rule-based Constraint Propagation: Theory and Applications*
- ILLC DS-2004-03: **Boudewijn de Bruin**  
*Explaining Games. On the Logic of Game Theoretic Explanations*
- ILLC DS-2005-01: **Balder David ten Cate**  
*Model theory for extended modal languages*
- ILLC DS-2005-02: **Willem-Jan van Hoeve**  
*Operations Research Techniques in Constraint Programming*
- ILLC DS-2005-03: **Rosja Mastop**  
*What can you do? Imperative mood in Semantic Theory*
- ILLC DS-2005-04: **Anna Pilatova**  
*A User's Guide to Proper names: Their Pragmatics and Semantics*
- ILLC DS-2005-05: **Sieuwert van Otterloo**  
*A Strategic Analysis of Multi-agent Protocols*
- ILLC DS-2006-01: **Troy Lee**  
*Kolmogorov complexity and formula size lower bounds*
- ILLC DS-2006-02: **Nick Bezhanishvili**  
*Lattices of intermediate and cylindric modal logics*
- ILLC DS-2006-03: **Clemens Kupke**  
*Finitary coalgebraic logics*

ILLC DS-2006-04: **Robert Špalek**

*Quantum Algorithms, Lower Bounds, and Time-Space Tradeoffs*

ILLC DS-2006-05: **Aline Honingh**

*The Origin and Well-Formedness of Tonal Pitch Structures*

ILLC DS-2006-06: **Merlijn Sevenster**

*Branches of imperfect information: logic, games, and computation*

ILLC DS-2006-07: **Marie Nilsenova**

*Rises and Falls. Studies in the Semantics and Pragmatics of Intonation*

ILLC DS-2006-08: **Darko Sarenac**

*Products of Topological Modal Logics*

ILLC DS-2007-01: **Rudi Cilibrasi**

*Statistical Inference Through Data Compression*

ILLC DS-2007-02: **Neta Spiro**

*What contributes to the perception of musical phrases in western classical music?*

ILLC DS-2007-03: **Darrin Hindsill**

*It's a Process and an Event: Perspectives in Event Semantics*

ILLC DS-2007-04: **Katrin Schulz**

*Minimal Models in Semantics and Pragmatics: Free Choice, Exhaustivity, and Conditionals*