Bericht Nr. 125, November 1997

Arbeitspapiere des
Sonderforschungsbereichs 340

ISSN 0947-6954/97

Sprachtheoretische Grundlagen für die Computerlinguistik

Robert van Rooy:

Attitudes and Changing Contexts

Universität Stuttgart Universität Tübingen
IBM Deutschland Informationssysteme GmbH

Arbeitspapiere des
Sonderforschungsbereichs 340

Sprachtheoretische Grundlagen für die ComputerLinguistik

Der Sonderforschungsbereich 340 ist eine Einrichtung der Universitäten Stuttgart und Tübingen in Kooperation mit der IBM Deutschland GmbH. Er wird gefördert durch die Deutsche Forschungsgemeinschaft. In den Arbeitspapieren werden der Fachöffentlichkeit in unregelmäßiger Folge aktuelle Arbeiten zugänglich gemacht. Die Arbeitspapiere sind—zum Teil auch on-line—zu beziehen über die Universität Stuttgart.

ISSN 0947-6954/97

Universität Stuttgart
Sonderforschungsbereich 340
Azenbergstraße 12
D-70174 Stuttgart

Universität Tübingen
Sonderforschungsbereich 340
Wilhelmstraße 113
D-72074 Tübingen

IBM Deutschland GmbH
Wissenschaftliches Zentrum
Inst. f. Wissensbasierte Systeme
Sonderforschungsbereich 340
Vangerowstraße 18
D-69115 Heidelberg

http://www.ims.uni-stuttgart.de/projekte/SFB340.html

Attitudes and Changing Contexts

Von der Fakultät Philosophie der Universität Stuttgart
zur Erlangung der Würde eines Doktors der Philosophie (Dr. phil.)
genehmigte Abhandlung

Vorgelegt von Robert van Rooy
aus Drunen, die Niederlande

Hauptberichter: HD Dr. Thomas E. Zimmermann
Mitberichter: Prof. Dr. Dr. h.c. Hans Kamp
Tag der mündlichen Prüfung:
28 oktober 1997

Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart
1997

Acknowledgements

This dissertation was written during the years that I was a member of the Graduiertenkolleg "Linguistische Grundlagen für die Sprachverarbeitung" at the Institut fur Maschinelle Sprachverarbeitung (I.M.S.) at the university of Stuttgart.

First of all I would like to thank Ede Zimmermann, my supervisor. Although he never forced me to work in any particular direction, this thesis would have looked like completely different without his influence. Not only did we discuss all issues relevant to this dissertation during my stay at the I.M.S., but he also critically read versions of all chapters of this thesis and made helpful suggestions with respect to both style and content. I was very fortunate having him as my supervisor.

I learned a lot from Hans Kamp during my stay at the I.M.S. through the courses he taught, his contribution to the several discussion groups on semantics and pragmatics we were both members of, and through the many critical but always insightful and stimulating questions and remarks he made after the several talks I gave in which he was in the audience. In an indirect way, the influence of Hans Kamp on the content of this thesis was considerable.

Paul Dekker read and discussed with me earlier versions of the first four chapters of this dissertation during his stay at the IMS in the beginning of 1997. This lead to considerable improvements on both style and content. I was happily surprised to see how close my analysis of especially indefinites and anaphora is to his more recent views on this subject.

I would like to thank Rob van der Sandt not only for teaching me all about the ins and outs of different theories of presuppositions during the time that I was a Philosophy student at the University of Nijmegen, but also for sending me to the I.M.S. at the time that I finished my M.A. thesis.

Uli Haas Spohn read and discussed with me an old version of the first chapter of this dissertation. I also thank her for allowing me to replace her as a member of the Forschungsgruppe "Logic in Philosophie" at the universities of Konstanz and Tübingen during the time that she has another appointment.

Special thanks to Ben Shaer who corrected the English of chapter 1, and a paper underlying chapter 3. He did a great job. I would also like to thank Graham Katz, who checked my English of the first part of chapter 2, and Peter Öhl, who translated the summary in German for me.

Lot's of thanks to all those who were working with me together in Herdweg the past years, especially, Dorit Abusch, Franz Beil, Steve Berman, Mariana Damova, Regina Eckardt, Tim Fernando, Arild Hestvik, Wolfram Hintzen, Peter Krause, Telmo Móia, Mats Rooth, Antje Roßdeutscher, Daniel Rossi, Ben Shaer, Renate Tanini, and Carl Vogel, and also to some others like Glenn Carroll, Michael Dorna, Annette Frank, Josef van Genabith, Jeroen Groenendijk, Manfred Kupffer, Nils van der Laan, Wolfgang Spohn, Arnim von Stechow, Henk Zeevat, and the members of the Graduiertenkolleg in Stuttgart, for valuable discussions, for the playing of billiards, cards, soccer, or table tennis, or at least for interesting gossip.

Thanks also to Daniel Heiming with whom I lived in a 'Wohngemeinschaft' in Stuttgart for more than three years.

Last, but not least, I want to thank my family who made it all possible, and who made me feel home at the times I was back in Holland.

Contents

Chapter 0

General introduction and overview

In the final chapter of his classic book about pragmatics, Gerald Gazdar (1979) wondered whether truth-conditional semantics can be autonomous with respect to pragmatics. Semantics is autonomous with respect to pragmatics, he said, when the truth conditions of natural language sentences can be determined without making reference to the presuppositions and other features of the contexts in which the sentences are used. At least since the rise of dynamic theories in the 1980s, wide agreement has been reached that this autonomy thesis should be given up: the truth conditions of an individual sentence cannot be determined without making reference to the facts about the conversational context in which this sentence is uttered. In this dissertation I will argue that what counts is not in the first place the preceding sentences of the discourse, but the *circumstances* in which these sentences are used, and the *attitudes* of the agents that use them. Although I believe that truth-conditional semantics can be fruitfully studied in abstraction from pragmatics, this does not mean that the beliefs and presuppositions of language users are irrelevant to the truth-conditional contents of natural language sentences.

In this dissertation I try to motivate and use a conception of meaning most explicitly defended by Stalnaker (1984). According to this conception, the meaning and content of linguistic expressions should be explained in terms of the intentions, beliefs and conventions of language users. This is in particular the case for modal discourse and for referential expressions: modal discourse should be explained in terms of beliefs and activities of rational agents; and the key notions of model-theoretic semantics, the notions of *reference* and *aboutness*, should be explained in terms of what speakers do by their use of a term, not by properties of the term itself. This conception of reference and aboutness is not only defended by Stalnaker, but was also the received view of semantics until the late 1960s. According to the traditional view, the expression (or thought) E refers to, or is about, R because (i) the speaker intends to refer to (or thinks about) the object (or set of objects) that satisfies the definite description the speaker or thinker associates with E (or is presupposed to be associated with E); and (ii) R satisfies this description. However, Donnellan, Kaplan, Kripke, Putnam and others have shown that the received view leads to counterintuitive results. On the basis of these observations it has sometimes been suggested that the denotation relation between terms and objects should be explained in causal terms independent of the intentions and beliefs of language users. I will follow Stampe (1977) and Stalnaker (1984), however, and argue that the causal theory of reference itself should be grounded in a (partly) causal, information-theoretic account of intentionality, and that therefore reference can be explained again in terms of beliefs and intentions of language users. Crucial to this account is that the content of a representational state need not be explained in terms of what has actually caused the representational mechanism being in the particular state it is, but might also be explained in terms of what under normal circumstances causes the mechanism to be in that state. In this way we can not only analyse singular terms like indexicals and proper names by a causal theory of reference, but general terms as well. Also, we can in this way analyse not only *knowledge* and perceptual attitudes in (partly) causal terms, but the more dispositional attitude of *belief* too. The main goal of chapter 1 will be to defend this picture, and to defend what I take to be a consequence of this picture: that from the believer's point of view a belief state should be modelled by a proposition individuated by truth conditions. I will defend this picture, and the consequence of it by arguing for a three-way strategy to solve some problems that arise on such an approach. First, I argue that we should make use of *diagonalisation* to account for the intuition that it might be unclear what, under normal conditions, is causally responsible for a certain representation. Second, I use a *counterpart theory* that allows for the possibility that agents have different representations of the same real object. Third, I will argue that many of the problematic aspects of attitude attributions are *pragmatic* problems, due to the extreme context-dependence of attitude attributions.

In chapter 1 I seek to reconstruct the Stalnakerian position with respect to content and belief attributions. It is partly built on the insight, due to Kaplan, Stalnaker and others, that it is good to make a conceptual distinction between two kinds of facts: (i) facts about the subject matter of thought or conversation, and (ii) facts about linguistic and speech conventions, and the conversational situation itself. This conceptual distinction will be used extensively in the following three chapters about anaphora and presuppositions.

In chapter 2 I will account for anaphoric relations across sentential boundaries on the basis of the intuition that pronouns are normally used referentially, and the assumption motivated in chapter 1 that referring is something done by speakers with their use of a term, not by the term itself: Which object is referred to depends on the intention of the speaker. Kripke (1977) taught us that a distinction must be made between *general* and *specific* intention. I will argue that for pronouns it is normally the specific intention that counts. Speakers normally refer back with their use of a pronoun, or short description, to the speaker's referent associated with the indefinite that figures as its syntactic antecedent. In this chapter I will show that by means of diagonalisation such an analysis can be pushed further than many have supposed; and that in fact this analysis is close to, but not identical with, modern theories of anaphora like Discourse Representation Theory (Kamp, 1981), File Change Semantics (Heim, 1982), and the more recent Dynamic Semantics due mostly to Groenendijk & Stokhof (1991). The reason will be that participants in a conversation are normally not only unclear about the facts relating to the subject matter of conversation, but also of some relevant facts about the conversational situation itself. Of course, sometimes a singular pronoun that takes an indefinite as its syntactic antecedent can be appropriately used although it does not refer to the specific speaker's referent of the indefinite. Sometimes it is only the general intention that counts. I will argue that to account for many of those cases we need descriptive pronouns in addition to referential pronouns. The former are pronouns that go proxy for a description recoverable from the sentence in which its syntactic antecedent occurs. In chapter 2 I will be concerned mainly with motivating this division of labour, implementing this analysis of pronouns into a dynamic theory of meaning, and using this two-tiered approach to account for phenomena problematic for the above-mentioned popular theories of anaphora.

In chapter 3 I discuss how to account for anaphoric relations across belief attributions, concentrating mainly on the problem of *intentional identity* made famous by Geach's Hob-Nob sentences. I will discuss how much of the popular view, which takes so-called unbound pronouns either as abbreviations for the antecedent clause or as variables bound by a dynamic existential quantifier, can be maintained. I will suggest that this view cannot be maintained. If we should account for Edelberg's (1986) *asymmetry problem* in semantics, I will argue that it is useful to take the notion of *speaker's reference* seriously in *semantics*.

What proposition is expressed by a sentence depends partly on the speaker's presuppositions. This is in particular the case for quantified statements like *Every German loves his Buick*. Intuitively, this sentence is only *about* the set of Germans with Buicks that the speaker presupposes is conversationally salient. If he is not allowed in a particular conversational context to presuppose that there is such a salient set of Buick-owning Germans, the assertion will be counted as inappropriate because it is not clear what proposition is expressed by the sentence. I believe that this is the intuition we have to capture, and I argue in chapter 4 that we can account for this intuition in a two-dimensional analysis of presuppositions in the style of Karttunen & Peters (1979). Any two-dimensional analysis of presuppositions is based on the assumption that there is a semantic value of any sentence that can be determined independently of the context in which it is used. Traditionally, this semantic value was supposed to be its truth conditions. However, this assumption gave rise to the so-called *binding problem*. Karttunen & Peters suggested that it is the assumption that we should represent what is asserted and what is presupposed by separate propositions or logical forms that is responsible for the binding problem. Indeed, it is now commonly assumed that presuppositions cannot be handled in a two-

dimensional theory of presuppositions. In this chapter I argue that this conclusion was drawn too quickly. Moreover, it is not in accordance with the conclusions we should draw from dynamic semantics. Traditionally, the semantic value of a sentence is its truth conditions, but according to dynamic theories it is its context change potential, a function from contexts to contexts. But once this insight is taken seriously in presupposition theory, it is again feasible to account for presuppositions in a two-dimensional way. In this chapter I argue that the binding problem can be solved in a two-dimensional theory by first determining the *context change potentials* of what is presupposed and what is asserted by a sentence independently of each other, and second, making the *truth conditions* of what is asserted dependent on what is presupposed.

According to the Lewis/Stalnaker analysis of conditionals, the truth conditions of a conditional sentence depends crucially on the speaker's intentions. The speaker's intentions, together with the antecedent and (other) facts about the actual world, select the relevant world(s) with respect to which the truth value of the consequent, and thus of the whole conditional, is evaluated. Stalnaker tried to make a stronger claim: the formal properties of the function that does this selection should be explained in terms of the beliefs and presuppositions of language users. He proposed that the analysis of conditionals should be related to the analysis of belief revision. I will discuss this project in chapter 5 and give some attention to Lewis's triviality result, which showed that what is expressed by a conditional must be even more context-dependent than the original Lewis/Stalnaker analysis suggested, if conditionals are to be explained in terms of conditional beliefs. I will argue in chapter 5 that most conditionals express propositions, but that the proposition expressed by an indicative conditional depends more directly on what is believed and presupposed by the speaker than the proposition expressed by a subjunctive conditional. Finally I will argue that what is expressed by a subjunctive conditional depends in such a systematic way on the conversational context that we could analyse them as variable strict conditionals, where the relevant accessibility relation is defined in terms of the Lewis/Stalnaker notion of 'similarity'. It differs from the traditional analysis in that the notion of 'closeness' can change its denotation through conversational means.

In the final chapter, I make use of the analyses of conditionals, belief revision, and rational decision discussed in the foregoing chapter in order to account for the meaning of some attitude verbs other than 'believe'. I will give the most attention to verbs of desire, and to the analysis of permission sentences. For the analysis of desire attitudes, I will argue that just like beliefs, also desires are closed under logical implication. Also just like beliefs, they are only closed under logical implication with respect to the *relevant* alternatives, but that it is very context dependent what the relevant alternatives are. With respect to permission sentences I will investigate in how far they can be accounted for in terms of possible world semantics. For some of those verbs I will also propose a way to account for anaphoric relations across attitude verbs.

Chapter 1

Belief and belief attribution

1.1 Introduction

According to the most straightforward account of belief attributions, the meaning of *a believes that A* in a particular context *c* is compositionally determined from the meaning of its parts in *c*. If it is assumed that meanings are assigned primarily to expressions, on this approach it seems that substitution puzzles like those that will be presented in §1.2 force one to assume that meanings and contents are really very fine-grained entities, and that a belief state should be modelled in a very fine-grained way. The problem with this approach is that it seems hard to give an *independent* motivation for such a fine-grained notion of content. The alternative strategy would be to start out by giving a philosophically motivated notion of content, independent of belief attributions. Such an independent notion of content will then typically be a rather coarse-grained notion. The fact that so many belief attributions still seem to be true and appropriate is then explained, according to this alternative strategy, partly in terms of the intentions and presuppositions of the agent who is making the belief attribution. It is the latter strategy that I will be defending in this chapter.

I will argue that the question *What is it that makes an attitude attribution true?* asks for a combination of a pragmatic account of intentionality defended by, for instance, Ramsey (1931) and a causal information-theoretic account as proposed by Stampe (1977) and by Dretske (1981). Both accounts motivate a rather coarse-grained analysis of belief states and of the content expressed by a sentence. On the basis of this I will argue that if we forget about the dynamics of belief and that of belief attributions, both of the above questions can and should be answered with: from the agent's point of view, by a set of possible worlds.

The main part of this chapter will address the question of to what extent the causal and the pragmatic accounts of intentionality are compatible with each other. The causal account sometimes seems to predict a too specific and sometimes a too unspecific notion of content and object of belief. I will discuss these problems mainly by looking at the traditional questions of how to handle *de dicto*, *de se* and *de re* belief attributions in a framework like Quantified Modal Logic. I will argue that most problems can be accounted for when (i) questions about *attitude attribution* are separated from questions about the *contents* of the attitudes themselves; and (ii) when we distinguish between an *object* and the body of *information* caused by this object.

In this chapter I start with the well-known substitution puzzle. Then I describe quantified modal logic in its most straightforward way. Afterwards I discuss two variants of what might be called the description theory of meaning, developed to solve problems that quantified modal logic faces, too. Kripke and others have raised strong arguments against this description theory of meaning, and have argued for an alternative causal theory of content. After I sketch the pragmatic account of intentionality, I will show that similar arguments used against the description theory of meaning also indicate that the pragmatic account of intentionality has to be supplemented by a causal, or information-theoretic, account. I argue that both the pragmatic and the causal information-theoretic accounts of content indicate that belief states looked at from the agent's point of view should be individuated by truth conditions, and thus should be represented by sets of possible worlds. Then I discuss some well-known problems raised by the assumption that causality plays such an important role in mental and linguistic representations. The remainder of this chapter is devoted to motivating and explaining a three-part solution for these problems. First, I argue that although the content of a mental or linguistic representation depends on external conditions, it might be unclear for believers, or for participants of a conversation, what the relevant external conditions are. Formally, I will argue that the meaning of an expression can be both index- and context-dependent, and that in a counterfactual

reference-context referential expressions might have a different referent than in the actual reference-context. I will argue that with the Stalnakerian technique of *diagonalisation* some problems concerning beliefs and belief attributions can be solved. I give special attention to self-locating beliefs, because they show the impossibility of a purely descriptive account of content, and thus are the greatest threat to a purely possible-world account of belief. Unfortunately, diagonalisation cannot solve all problematic belief attributions. First, it cannot account for the fact that in different conversational contexts the same attitude attribution may have a different truth value although the agent himself has not changed his mind. Second, it cannot solve those problems for a formal theory that wants to take seriously the issues that Kripke and others have raised for *de re* belief attributions where individuals are used to characterise a belief state. To account for the latter, I will argue that we need some kind of counterpart theory that allows for the possibility that, for instance, one individual in one world has two distinct representatives in another, and that this is compatible with an account of content that is not purely descriptive. Such a counterpart theory is the second part of the strategy I want to defend. But perhaps the most essential part of the strategy is an account of the extremely context-dependent nature of belief attributions. I will argue that the *pragmatics* of attitude attributions is very complicated: not only is what is expressed by a belief attribution dependent on the intentions and presuppositions of the attributer, but also how we should represent what the agent believes. In the last part I formulate a double indexing counterpart semantics for modal logic, where I try to account in a formal way for most of the ideas argued for in this chapter. Unfortunately, I am only partly able to account in a formal and systematic way for the *pragmatics* of belief attributions. But the formalism I propose will at least be compatible with such a systematic pragmatic account.

1.2 The substitution problem

On a naive understanding of definite noun phrases, like *the author of Waverly* and *the number of planets*, the meaning of those terms can be equated with their actual referents. But if this is assumed together with what is called Frege's *principle of compositionality*, which says that the meaning of a complex expression is a function of the meanings of its direct parts, we can derive the following *substitution principle* for definite expressions:

Two definite expressions T and T' with the same referent (in the same context of use) are interchangeable with each other in that particular context without change in meaning of the sentence in which they occur.[1]

However, as Frege (1892) noted, this principle leaves something unexplained: if the two sentences $T = T$ and $T = T'$ have the same truth value, how can it then be explained that the first is trivially true and known *a priori*, while the second is informative and *a posteriori* ?

The problem is even worse when we substitute co-referential definite terms for each other when these terms are used in sentences embedded under expressions like *George IV believes that* and *It is necessary that*. On the basis of the above substitution principle, we predict that from (1a) and (1b) we can derive (1c), and from (2a) and (2b) we can derive (2c):

(1a) George IV believes that Sir Walter Scott is Sir Walter Scott.
(2a) It is necessary that (9 > 7).

(1b) The author of *Waverly* is Sir Walter Scott.
(2b) The number of major planets is 9.

(1c) George IV believes that the author of *Waverly* is Sir Walter Scott.
(2c) It is necessary that (the number of major planets > 7).

[1] The substitution principle can be stated in more general terms, but we will not need this here.

But, as noted by Frege (1892), Russell (1905) and Quine (1953), these inferences do not go through; although the (a) and (b) sentences are true in the actual world, the (c) sentences are not. It is only natural to assume that whenever two sentences uttered in a certain context differ in truth value, the meanings of the two sentences must be different in this context, too. The meaning of a sentence is compositionally determined form the meanings of its parts. It follows that if two sentences have different truth values, but only differ from each other in that the definite term T in the first sentence is replaced by definite term T' in the other sentence, the meanings of the two definite terms cannot be the same in that context of use. The question that arises is how we should account for the difference in meaning of the two definite terms that are co-referential in a particular context. I will discuss this question in terms of the framework of quantified modal logic. As we will see, the substitution problem rears its ugly head in this framework in exactly the same way as we saw above.

1.3 Quantified modal logic

It is well known that combining propositional modal logic with first-order predicate logic is not as obvious as it seems. It seems that some principles that are crucial for first-order predicate logic have to be given up if combined with propositional modal logic. Quantified modal logic (for a formal language L) can be stated as follows:

The syntax for the language L need not be specified: it is quite straightforward and will be implicit in the semantic clauses given later. It is only crucial that we make the assumption that all definite noun phrases are represented as individual constants. Pointed models, M, are septuples $<W, w_0, D, A, R, K, I>$, where W is a non-empty set of *worlds*; w_0 a designated element of W, representing the actual world; D a non-empty set of *objects*, the domain; A a set of *agents*, a subset of D; R a binary relation on W; K a function in $[(A \times W) \rightarrow \wp(W)]$; and I the interpretation function. Individual constants will be interpreted as individuals, elements of D. In the most straightforward way it is assumed that there is one fixed domain over which quantifiers range, and that if a variable x refers to d in one world, it refers to d in all possible worlds in which d exists. For simplicity I will assume that all individuals exist in all possible worlds. The interpretation function I meets the following conditions:

- for each constant c: $I(c) \in D$
- for each n-ary predicate symbol P of L and each $w \in W$: $I_w(P) \subseteq D^n$

$[[t]]^{w,g}$ = I(t), if t is a constant symbol
= g(t), if t is a variable.

The satisfaction conditions are then defined as follows (where we leave out the superscript for the model):

$w,g \models P(t_1,...,t_n)$ iff $<[[t_1]]^{w,g},..., [[t_n]]^{w,g}> \in I_w(P)$
$w,g \models t_1 = t_2$ iff $[[t_1]]^{w,g} = [[t_2]]^{w,g}$
$w,g \models \sim A$ iff $w,g \not\models A$
$w,g \models A \wedge B$ iff $w,g \models A$ and $w,g \models B$
$w,g \models \forall xA$ iff for all $d \in D$: $w,g[x/d] \models A$
$w,g \models \Box A$ iff $\forall w' \in R(w)$: $w',g \models A$
$w,g \models Bel(t, A)$ iff $\forall w' \in K([[t]]^{w,g},w)$: $w',g \models A$

Truth and validity are defined in the usual way. We say that A is *true* in M, iff for all g, $w_0,g \models A$ in M, and A is *valid* iff it is true in all models.

If '$\exists xA$' is an abbreviation for '$\sim\forall x\sim A$', in this logic the following principles are valid:

(3)	$\forall x,y[x = y \rightarrow (A(x) \leftrightarrow A(y))]$	Substitution of Identicals	(SI)
(4a)	$A(t) \rightarrow \exists xA(x)$	Existential Generalisation	(EG)
(4b)	$\forall xA(x) \rightarrow A(t)$, for all t	Universal Instantiation	(UI)[2]

The reason is that the interpretation function assigns to each individual constant of the language one single individual in D. By the truth conditions for quantified sentences variables are also analysed in this way.

But the assumption that these formulae are valid leads us back to the substitution puzzles in modal and attitude contexts, as discussed above. The reason is that by (SI) and (UI) we can derive the principle that any two co-referential singular terms can be substituted for each other without change of truth value. By (UI) we can first derive $\forall x[t = x \rightarrow (A(t) \leftrightarrow A(x))]$ from (SI); and then by (UI) again, $t = t' \rightarrow (A(t) \leftrightarrow A(t'))$. This 'theorem', however, leads to problems if we assume, as Frege did, that proper names, indexicals and definite descriptions should be treated as singular terms, and logically represented as individual constants.

The assumption of Frege's hypothesis in our logic above not only leads to substitution puzzles; existential generalisation is also problematic. The following formula will be valid: $\Box(a = a) \rightarrow \exists x[\Box(x = a)]$. But if *a* stands for *the morning star*, it seems that the formula should not be valid. Intuitively it does not follow that by accepting the antecedent of the above formula, we are committed to the view that there is an object that is necessarily the morning star. According to Quine (1943), objects don't have essential properties, contrary to what quantified modal logic predicts. For similar reasons, negative existential sentences like *The difference between A and B does not exist* (Russell, 1905) are also problematic, if in each world there are only existing objects.

There are two ways around these problems. One can concentrate on the semantic value of the *term* or on the way the relevant agents think of the *referent* of the term. By adopting the first strategy it is assumed that the semantic value of the term is world-dependent in some way or another. It seems natural that a belief attribution is true iff the agent would assent to the embedded sentence of the attribution. Following the second strategy, however, this assumption is given up. Of course, a mixed strategy would be to use both. For the moment I will first concentrate on the first kind of approach.

1.3.1 Individual Concepts

Church (1943) and Carnap (1947) proposed to account for the substitution problems by following Frege (1892). Accordingly, singular terms are not interpreted as individuals, but as individual concepts instead. An individual concept is a function from worlds to individuals, an intension. The terms refer only indirectly to actual objects, via the intensions. Once individual terms are interpreted by individual concepts, it is only natural to assume that two individual terms are interchangeable with each other in a certain context without change of truth value if the concepts by which they are interpreted refer to the same objects in all relevant possible worlds. Although two individual concepts may refer to the same individual in this world, they don't have to do so in all relevant possible worlds. Thus, their identity conditions are very strong. In this way the substitution problem can be solved. For instance, in another world the noun phrase *the number of planets* does not have to refer to the number 9. Moreover, individual concepts don't have to be total functions; it

---

[2] Note that EG and UI are two aspects of the same principle; we can derive one from the other by contraposition and double negation elimination.

might be that a concept has no instantiation in the actual world. In this way, negative existential sentences can also be accounted for. Logics where the presupposition is given up that singular terms (variables and constants) always designate an existing object are called free logics.

On the assumption that all individual terms should be treated in the same way, it follows that individual variables, a subset of the individual terms, should also range over entities that can refer in such a world-dependent way. Variables should not range over real-world objects, but over individual concepts instead.[3] By assuming that all individual terms refer to individual concepts, the free logic variants of Universal Instantiation (FUI) and Existential Generalisation (FEG) can still be assumed. But this is really problematic. The problem is that if variables range over all individual concepts,[4] the following two principles become valid (if ιyB is read as *the y such that B*, if A[y/x] stands for A with y substituted for x, and E is the existence predicate):

(a) $\forall xA \rightarrow (E(\iota yB) \rightarrow A[\iota yB/x])$ (FUI)

(b) $(E(\iota yB) \wedge A[\iota yB/x]) \rightarrow \exists xA$ (FEG)

If A is true for all individual concepts, it is certainly true for the special concept ιyB, and if A is true for the special concept ιyB, there is at least one individual concept for which A is true. However, these principles should not be valid if the logical system is used to formalise natural language. Consider the following examples (Gamut, 1991, Vol. 2, p. 59):

(5a) Everybody can lose this game.
(5b) The winner can lose this game.

(6a) The president of the USA has to be born in the USA.
(6b) There is somebody who has to be born in the USA.

The problem is that we have also quantified over function-words like *the winner* or *the president of the USA*. We quantify over too many individual concepts. Quantifying over all individual concepts does not correspond with the suggestive case of Priorian tense logic.[5,6]

1.3.2 The description theory

Whereas Church and Carnap tried to save modal logic by following Frege in intensionalising the interpretation of terms, Smullyan (1948) tried to save modal logic by following Russell (1905) in giving up the Fregean Hypothesis that all definite noun phrases should be analysed as individual constants. They should be eliminated in favour of predicates whose meanings are world dependent. A sentence like *The difference between A and B does not exist* is only an apparent counterexample to existential generalisation because the phrase *the difference between A and B* is not really a singular term. The sentence is translated as something like $\neg\exists x\exists y[\text{Diff-A-B}(x) \wedge \forall z[\text{Diff-A-B}(z) \rightarrow z = x] \wedge x = y]$, and existential generalisation does not apply. The equations (1b) and (2b) can be true

---

[3] Carnap (1947) assumed that variables denote both senses (concepts) and denotations.

[4] Something that is not argued for by Garson (1984).

[5] If it is assumed that in all worlds we always quantify over the same set of individual concepts, we also face another obvious problem: a *de re* formulae like $\exists x\Delta Px$ will have the same truth condition as the dicto formula $\Delta\exists xPx$.

[6] Treating descriptions as world dependent singular terms does not force one to assume that all terms should be treated in this way. Thomason & Stalnaker (1968), for instance, treat descriptions as world dependent singular terms, but let the quantifiers range only over rigid individual concepts. As a result, (UI) is no longer valid.

in the actual world, without being true in all worlds, or without being true in the belief worlds of George IV. In this way the inferences from (a) and (b) to (c) are blocked. Modal logic makes sense if we are prepared to look at certain (apparent) singular terms as abbreviations for more complex descriptions. According to Russell's theory of descriptions, sentences in which descriptions and intentional constructions occur are ambiguous. (2c), for instance, is ambiguous between the logical paraphrases (2d) and (2e):

(2d) $\Box\exists x[\forall y[\text{Number-of-planets}(y) \leftrightarrow y = x] \wedge (x > 7)]$

(2e) $\exists x[\forall y[\text{Number-of-planets}(y) \leftrightarrow y = x] \wedge \Box(x > 7)]$

The difference between (2d) and (2e) is that the description *the number of planets* is evaluated in different worlds. The Russellian distinction of scope can account for the fact that we cannot infer (2c) from (2a) and (2b), because (2d) doesn't follow from the (a) and (b) clauses.[7]

The Russellian tactic works quite straightforwardly for definite terms like *the author of Waverly* and *the number of planets*. But to account for all substitution puzzles and problems concerning negative existential sentences raised by definite noun phrases, this tactic also requires that proper names and even most indexicals should not be represented as individual constants. These definite expressions must also be eliminated, according to Russell, in favour of predicates. Thus, whether we follow the Fregean strategy and assume that singular terms can refer to different entities in different worlds (because there is a set of properties associated with the singular term); or we eliminate (apparent) singular terms in favour of predicates, as proposed by Russell, we still have to associate with such natural language expressions sets of properties or descriptions that uniquely determine their referents.[8] These sets of properties must then be equated with the intensions of the expressions. Let's call this *the FR description theory*. Traditionally this FR description theory is assumed not only for singular noun phrases, but for natural language expressions in general. The intension of an expression is then taken to be a set of properties that uniquely determines its extension.

1.4 Problems for the description theory of reference and externalism

We have seen that the most straightforward solution to the substitution puzzles is to assume that some set of descriptions or properties is associated with a proper name, *N*, and determines the referent. We have also assumed that a speaker refers on a particular occasion to *a* by using *N*, when *a* is, or is believed to be, the unique individual that has all or most properties associated with *N*. If there is no such unique individual, the name does not refer. Thus, we assumed that for a speaker to refer with *N* to *a*, it must be the case that the speaker associates with *N* a set of general properties or descriptions, and that *a* is the unique object that fits these general descriptions best. However natural this *description theory of speaker's denotation* might be, Donnellan (1970) and Kripke (1972) have shown that it leads to counterintuitive results. Kripke argued that uniquely fitting some set of descriptions that the speaker associates with a proper name is neither a necessary nor a sufficient condition for a successful use of it. It is *not necessary* that the *speaker* has an identifying set of descriptions in mind for the successful use of a proper name, because ordinary people can, for instance, use the name *Feynman* to denote the physicist Feynman even though they have no uniquely identifying set of descriptions in mind. To uniquely satisfy all or most of the descriptions associated with a proper name is also *not* a *sufficient*

---

[7] Of course, quantified modal logic as stated above is committed to the *meaningfulness* of formulae like (2e). Quine concluded that as a result it is committed to a suspect sort of Aristotelian essentialism.

[8] It should be noted, though, that whether we rely on individual concepts or use the description theory as a formal tool, we are not committed to the view that all definite noun phrases refer to objects because this object is the unique object that fits best with the description (or set of descriptions) associated with the definite noun phrase. For a recent use of the description theory without adopting the Frege-Russell account of determining reference, see Muskens (1989).

condition for an individual to be referred to by the name. This point is made clear by Kripke's Gödel example. If someone associated with the name *Gödel* only the description *prover of the incompleteness of arithmetic* he would still denote Gödel and be saying something false of him in uttering *Gödel proved the incompleteness of arithmetic* if somebody different from Gödel was the actual prover of what is known as 'Gödel's incompleteness theorem'. Besides giving similar kinds of counterexamples to the description theory of speaker's denotation, Donnellan (1970) also pointed out that what is referred to by a proper name by a speaker on a particular occasion depends not only on the intention of the speaker, but also on the conversational context. Consider a student who is known to be acquainted with two different persons with the name *J.L. Aston-Martin*. The one is a famous philosopher that he knows from reading his books, and the other a non-famous person that he knows because he met him at a party the night before. Unfortunately, the student wrongly assumes that both persons are one and the same, that he is acquainted with one person named *J.L. Aston-Martin* in two different ways. He associates with the name a set of descriptions that does not uniquely fit one individual: some descriptions fit the famous philosopher, while some others the man he met at the party. Still, Donnellan argues, in some conversational situations the student will unambiguously refer to the famous philosopher by his use of the name *J.L. Aston Martin*, while in others he will unambiguously refer to the man he met at the party. Obviously, this is something the description theory of speaker's denotation cannot account for.

Perhaps the extension of a proper name does not depend so much on the descriptions the speaker associates with it, but on the set of descriptions most people in the relevant linguistic community associate with it. It is then this set of descriptions that determines the reference. However, Donellan and Kripke showed that this, too, cannot be the case. Kripke's example of Gödel, for instance, shows that this has counterintuitive results. And as Donnellan and Kripke observed, if we associate with the name *Aristotle* the description *the teacher of Alexander*, it would also lead to the conclusion that the statement *Aristotle was the teacher of Alexander* is true solely because of the meaning of the proper name. This again seems counterintuitive.[9]

Kaplan (1989) and Perry (1977, 1979) have argued against description theories of reference of *indexicals* (pure and demonstrative) on grounds very similar to those presented by Kripke and Donnellan. There seems to be no plausible candidate that could be the speaker's meaning of an indexical like *today*. First, it cannot be the description the speaker associates with the relevant day. For suppose that the description I associate with 7 October picks out 8 October, instead of 7 October, because I have taken a long nap. In that case, we would predict that the proposition expressed by my utterance of *Today the weather is fine* on 7 October was that on 8 October the weather is fine. Clearly, this is the wrong prediction. Second, suppose that the meaning of *today* is what a competent speaker of English associates with the word *today*. In that case the intension of *today* does not change from day to day. But this should be the case if the intension of the sentence is compositionally determined by the intensions of its parts, and the proposition expressed by it depends on the day of utterance. Similar arguments can be given against a description-account of pronouns demonstratively used.

By very much the same kind of arguments, Kripke, Putnam and Burge convincingly argued that the set of properties that speakers or agents associate with *common nouns* should also not be equated with the meaning of the noun. First, the meaning cannot be the description the *speaker* associates with the term. This is made very clear by the 'twin earth' stories given by Putnam (1975) and Burge (1979). These stories always involve a comparison between two almost identical persons (twins): one in the actual world and one in a counterfactual world minimally different from the actual world. In Putnam's story, the

---

[9] Some have argued that a set of descriptions fits a unique individual, if this individual is the unique individual that fits most descriptions of this set. Donnellan and Kripke also convincingly argued against such a weaker variant of the description theory.

stuff that the inhabitants of the counterfactual situation call *water* is superficially the same as the stuff *we* call *water*, but its chemical structure is not H_2O, but XYZ. If, then, both the earthling and his twin assert *Water is the best drink for quenching thirst*, intuitively they have said something different. But how can this be if they associate exactly the same set of properties with the word and if speaker's description determines reference? A similar 'twin earth' story invented by Burge (1979) shows that the problem is caused not only by a very limited set of terms. For almost all expressions, stories can be invented showing that it is not the description that the speaker associates with an expression that determines its extension. The reason is that linguistic practices of members of the agent's community are crucial in determining the extension of a term. Perhaps then what counts are the properties associated with the term by most speakers, or the relevant specialists of a linguistic community. But Putnam shows that also this cannot be the case for natural kind terms. The same 'twin earth' story is told, but is now situated in 1750. Specialists on earth and twin earth are not yet able to see any difference between H_2O and XYZ. But intuitively, even if a typical Twin-earthian (twin-) English speaker utters *Water is the best drink for quenching thirst* on earth, he is not talking about H_2O.

On the basis of these arguments Kripke and Putnam claim that the meaning of at least proper names and natural kind terms is not the set of descriptions associated with them, but simply the entity or stuff that is the *source* of the reference-preserving link from the initial baptism of the name to the speaker's use of the name. However, the existence of a causal link by itself is not enough, since it leaves out an important *intentional* element. It should at least also be the case that the speaker intends to use the name in the same way as it was transmitted to him via other members of the community. Evans (1973) argued that a causal link is necessary, but this causal link should not be the link between the initial dubbing of the name to the speaker's current use of the name, but between the body of information relevant to the speaker's use of the proper name on a particular occasion and the object that is the dominant causal origin or source of this body of information.[10] The causal account of reference can be extended to other referentially-used expressions. It could be the case that two objects are the source of a particular body of information relevant to the speaker's use of a referential expression on a particular occasion. In those cases, according to Evans, we say that on that occasion the speaker refers with the expression to the object that is the *dominant* source of this body of information. By making the referent of a name dependent on the dominant source of the relevant body of information, Evans can account for the fact that names have changed their denotation (like *Madagascar*). Consider the following example:

> If it turns out that an impersonator had taken over Napoleon's role from 1814 onwards (post Elba) the cluster of the typical historian would still be dominantly of the man responsible for the earlier exploits and we would say that they had false beliefs about who fought at Waterloo. If however the switch had occurred earlier, it being an unknown Army officer being impersonated, then their information would be dominantly of the later man. They did not have false beliefs about who was the general at Waterloo, but rather false beliefs about that general's early career. (Evans, 1973)

With a referentially-used expression we refer to the dominant source of the information 'responsible' for that use of the expression on a certain occasion. For a proper name (or any other definite term) however, it is not primarily the *speaker's* body of information that counts. For a speaker to refer to a particular entity on a particular occasion, it is not enough that this entity is the dominant source of the information of the *speaker* that is relevant for his use of the proper name.[11] For *N* to be a name for *a* it should also be the case that there

---

[10] For more on the distinction between a purely causal account and an informational account, see Dretske (1981, ch. 1). Kripke (1980, addenda (e)) argued that Evans was really a proponent of the description theory after all, by his use of the notion 'information'. At least since Dretske's book it should be clear that Evans was not.

[11] But it is a necessary condition for a speaker to refer with *N* to individual or stuff *a* that *a* is the dominant source of the content of *his thoughts* relevant to his use of *N*. How else to account for the intuition that when a typical twin-earthian (twin-) English speaker utters *Water is the best drink for quenching thirst* on

exists a convention among the speakers of the relevant linguistic community that *N* could be used to refer to *a*. This convention brings in the *social element* of the meaning of a proper name, and is to be explained in terms of beliefs and intentions of the members of the community. In sum, according to Evans' informational account of proper names, the information associated with a proper name plays its part, although the causal link is necessary. Since this causal element is still part of the analysis, *a* is not the referent of proper name *N* because *a* fits best with the information associated with *N*, but because it is the dominant source of this body of information. An object can be the dominant source of a particular body of information even if it does not fit this information very well. It follows that if *P* is one of the properties we associate with *N*, we still do not know that the sentence *N is P* is true *a priori*.

If the reference of an expression on a particular context of use depends on the causal link between this reference and a body of information, and if the relevant body of information must sometimes be determined in a social way, reference is determined by *externalist* means. Another aspect of externalism is needed to account for the fact that the student Donnellan (1970) talks about can refer to two different individuals by his use of *J.L. Aston Martin* in different conversational contexts. In the two different conversational contexts that *we* are in, there are different bodies of information that the student has which are relevant, and these have different sources.

It is pretty clear how a causal theory of reference can determine the content of a proper name. But how can it determine the content of a natural kind term like *water* or predicates like *arthritis* or *red*? It seems reasonable that according to the causal externalist account of content, the content of such expressions is determined in terms of a conception of normal or optimal conditions. With our use of the expression *water* we refer to H_2O because we want to use this term to refer to stuff that has certain observable properties, and if conditions are normal or optimal it is only H_2O that has these properties. We can account for the fact that a typical Twin-earthian (twin-) English speaker does not refer to H_2O by his use of *water*, because the normality or optimality conditions in terms of which content is determined are contingent ones (see Stalnaker, 1993). On twin earth it is not H_2O but XYZ that has the relevant observable properties, and is 'responsible' for the use of their term *water* by twin-earth (twin-) English speakers. But then, why is it that when water is H_2O, it is also *necessarily* H_2O? The reason is that the notion of normal or optimal conditions is a modal notion. Once we have found out the optimality conditions in the actual world with respect to the word *water*, these conditions determine a set of worlds in which they hold. In our case we can say that water is necessarily H_2O, because for necessity we look only at worlds in which the optimality conditions in the actual world with respect to the word *water* hold (compare Van Fraassen, 1977).

If we say that with respect to alethic modalities expressed by adverbs like *possibly* and *necessarily* we consider only worlds where the ideal conditions for the actual world hold, we can understand why Kripke argued that if we know what the actual extension of *E* is, the question of what *E* refers to in the relevant counterfactual possible worlds does not come up. We consider only such counterfactual possible worlds that has a unique primitive counterpart of the actual denotation of *E* , because the ideal conditions of the actual world also hold there. It follows that if a statement that asserts the identity of *E* and *E'* is true, the two terms also have the same extension in every metaphysically possible situation considered, and thus have necessarily the same extension. Consider the case of individuals. If proper name *N* refers to *a* in the actual world, we don't have to look at properties this individual has in common with another individual *b* in another world to decide whether they are counterparts of each other or not, that is, whether *N* refers in this other world to *b*.

It is possible that the individual with the same *thisness* as *a* has properties in such a counterfactual world, completely different from those that *a* has in the actual world.[12]

1.5 The pragmatic account of intentionality

The problem of intentionality is to explain how certain things can represent, be about, or directed to, other things. A number of things have this representational capacity; natural language and minds or attitude states of agents are two prominent examples among these. According to one account of intentionality, the representational capacity of language and other media should be explained in terms of the intentionality of the attitude states of agents. So, how can an attitude state represent something 'outside of itself'? The usual (holistic) strategy for answering this question is to say that this representation relation has to be explained in terms of relations to propositions. An agent represents or is directed towards another physical object if he/she stands in a certain attitude relation to the proposition that involves or is defined in terms of this other physical object. But then, what is it to stand in a certain attitude relation to a proposition? Propositions are abstract objects, and how can we stand in a relation at all to such an abstract object? Measurement theory suggests an answer: an agent stands in a relation to an abstract object if by this object we can measure the state of the agent. For a dispositional predicate like *believe*, this measure is usually stated in terms of counterfactual relations:[13] Agent *x* stand in a relation *R* to the proposition *P*, R(x, P), if there is a certain predicate *F* such that the following counterfactual is true: if it were the case that *P*, then *x* would be *F* (Stalnaker, 1984, 1994). Because *F* is a predicate saying something about how *x* stands in the world, a true attitude attribution characterises an individual as being in a certain state by saying something about the relation the agent bears to the world.[14]

According to a purely *pragmatic* (and almost behavioristic) account of intentionality, as defended by, for instance, Ramsey, 1931; Dennett, 1969; Stalnaker, 1972; and Lewis, 1974, the *R* is explained in terms of something like a *tendency to bring about* relation. An attitude of an agent is about, or directed to, an object or state of affairs because the agent is disposed to perform actions which involve this object or state of affairs. The attitude of the agent is this dispositional state. Proponents of this account assume that attitude attributions are normally made to explain behaviour. A person performed a certain action, because by doing so he could satisfy his desires in a world in which his beliefs are true. For this to work, it has to be presupposed that the behaviour of the agents can be rationally understood and that attitudes, in particular belief and desire, are correlative dispositions, or functional states, of such a rational agent. These states are individuated by the role they play in determining the behaviour of the agent who is in such a state. For instance, if *R* is *desire*, for an agent to desire that *P* means that the agent is disposed to perform actions that tend to bring about *P* in a world in which his beliefs were true. Analogously, if *R* is *belief*, the agent believes that *P* if he is disposed to perform actions which tend to satisfy his desires in worlds in which *P* and his other beliefs were true (Stalnaker, 1984, p. 15). A belief/desire system is correctly attributed to an agent if the actions that the agent performs can be explained in terms of this belief/desire system (and a theory of rational behaviour). And if

---

when a typical twin-earthian (twin-) English speaker utters *Water is the best drink for quenching thirst* on Earth, he is not talking about H_2O?

[12] It is normally assumed that if individual *a* in *w* has the same *thisness* as *b* in *w'*, *a* and *b* are really the same individual and that this one individual lives in different possible worlds. Normally I will do as if this is the case, but officially I don't want to be committed to this. All I will assume is that two individuals in two different possible worlds have the same *thisness* if the two are counterparts of each other, where the counterpart relation is not defined in terms of qualitative similarity. See the later discussion of counterpart theory why, I think, this is not incompatible with the views of Kripke (1972).

[13] By 'counterfactual' I mean here (and in the rest of this dissertation) the kinds of conditionals that are analysed in Lewis (1973).

[14] I am not sure whether the Stalnakerian approach I will defend can give a fully naturalistic account of intentionality in this way. What it certainly will do, though, is to explain notions that are thought to be mysterious, like *content* and *intentionality*, in terms of notions that are considered not to be so problematic.

the actions of the agent that involve object *a* can be explained in terms of the intentional state, the intentional state of an agent can be said to be about object *a*. So, in a manner similar to the FR theory of description, aboutness is determined by *fit* with reality.

The pragmatic picture suggests that propositions, the objects of attitudes, should be modelled by sets of possibilities. Why? The reason is that rational agents are seen as *deliberators*. A deliberator is an agent who considers various possible actions and determines his choice by his beliefs about the possible outcomes of these alternative actions and by the desirability of these possible outcomes. This picture of rational activity suggests that the primary objects of attitudes are sets of alternative outcomes of possible actions, or alternative ways that the world might be (Stalnaker, 1984, p. 4). These possible outcomes of actions can be thought of as possibilities that are maximally specific with respect to all of the issues relevant in the deliberation. Thus, if we want to say that an attitude state like *belief* is modelled by a set of possibilities, these possibilities are only as fine-grained as demanded by the conversational context.

Not only should belief states be modelled by sets of possibilities, but this all kinds of acceptance attitudes of the agent - in particular, the propositional attitude of *presupposition* (Stalnaker, 1973, 1974) - should be also. Just as beliefs and desires are functional states of rational agents, so too are presuppositions. All these attitude states belong to a theory of rationality, which is required to explain why certain behaviour of rational agents is appropriate when it is (Stalnaker, 1970b). The attitude of presupposition is needed to explain the agent's behaviour when he is engaged in a conversation. The appropriateness of his communicative actions is to be explained in terms of what he presupposes.

Saying that the objects of attitudes are sets of possibilities does not necessarily mean that a belief state of an agent should be represented by one set of possibilities. It might be that various thoughts the agent has are not integrated. In these cases he doesn't have a single coherent conception of the world. A belief state should not be modelled by a set of possibilities, but rather by a set of such sets. Each compartment can be thought of as what he believes if a certain question were asked or a certain problem posed. Thus, what somebody believes can be thought of as a function from problems to sets of possibilities. This latter set of possibilities is then what the agent *explicitly* believes, how he is disposed to act, in a situation in which the problem is posed. What somebody *implicitly* believes might be thought of as union of the set of compartments. Although implicit beliefs are closed under logical consequence, for explicit beliefs this is the case only with respect to each compartment separately. Stalnaker (1984, ch. 5) suggests that deductive reasoning might be thought of as trying to integrate different compartments of one's belief state with each other.[15] The beliefs of agents are not closed under deduction because not every compartment is always accessible. Questioning helps to make explicit what was only implicitly believed before (Stalnaker, 1991).

1.6 The causal information-theoretic account of intentionality

Although the account of intentionality sketched above is essentially an externalist one, the arguments of Kripke, Putnam, and Burge that are problematic for the description theory of meaning are also a threat to the purely pragmatic account of intentionality.[16] The pragmatic account of intentionality alone leaves the content of belief underdetermined. It cannot explain why Oscar (one of the twins in Putnam's twin earth example) is *thinking about* H_2O if he is thirsty and asks *Can someone give me some water?* According to a purely pragmatic account of intentionality, he might as well be thinking about XYZ. But we don't need these artificial twin earth stories to see that the pragmatic account by itself cannot solve the problem of intentionality. Just like for the description theory, *fit* is not enough.

---

[15] See also the neighbourhood semantics of Montague (1974), and the cluster models of Fagin & Halpern (1988).

[16] Although they were in the first place directed only against individualistic accounts of intentionality.

What makes an assignment of a system of belief and desire to a subject correct cannot just be that his behaviour and behavioural dispositions fit it by serving the assigned desire according to the assigned beliefs. The problem is that fit is too easy. The same behaviour that fits a decent, reasonable system of belief and desire also will serve countless very peculiar systems. Start with a reasonable system, the one that is in fact correct; twist the system of belief so that the subject's alleged class of doxastic alternatives is some gruesome gerrymander; twist the system of desire in a countervailing way; and the subject's behaviour will fit the perverse and incorrect assignment exactly as well as it fits the reasonable and correct one. Thus constitutive principles of fit which impute a measure of instrumental rationality leave the content of belief radically underdetermined. (Lewis, 1986, p. 38)[17]

Just like the arguments of Kripke and Putnam motivated a causal theory of reference, the twin earth stories of Putnam and Burge motivated a causal theory of intentionality. Some have concluded from Putnam's twin earth story that both the pragmatic account of intentionality and the description theory of meaning are generally satisfactory, their problems being limited to only a narrow range of cases. However, as Burge's (1979) extension of Putnam's twin earth story made clear, the problem for the above theories is much more general. This suggests that the notion of intentionality should, at least partially, be analysed causal information-theoretic terms. How should this causal information-theoretic account be cashed out?

I believe that this should be done in terms of counterfactual relations. Remember that the relation between agent x and abstract proposition P was in general defined in terms of a counterfactual definition of R(x, P). The pragmatic picture explains this relation in terms of a *tendency to bring about*. According to the information-theoretic account defended by Stampe (1977) and Stalnaker (1984), this relation is analysed in terms of a *tendency to carry information*, a relation of *indication*. For a certain mechanism to be a representational mechanism about a certain environment, the mechanism must be able to be in various alternative states that tend to vary systematically with variations in the environment. When this mechanism tends to be in state R when P is the case, this mechanism's being in state R can be said to contain the information that P; that is, it indicates that P is the case. Normally, the internal state of a representational mechanism tends to vary systematically with the states of the environment, and thus indicates something about the environment, because the mechanism's being in a certain internal state is *caused* by this environment. If this is the right way to explain why mechanisms can represent something outside themselves, and thus have content, it suggests that the notion of content, or the indication relation, should be analysed in terms of nested counterfactual conditionals: *if* conditions are normal or optimal, and *if* various alternatives to P were true, then the believer would be in various alternative states (Stampe, 1977, and Stalnaker, 1988). To the normal conditions, fidelity conditions (Stampe, 1977), or channel conditions (Dretske, 1981), belong both conditions external to the agent and conditions related to the internal functioning of the representational mechanism. The reason for this use of nested counterfactuals should be obvious: if conditions are not normal, P might hold without the internal mechanism being in state R, and it is the relevant alternative states that the environment could be in that determine the content of the internal state.

If content is explained in this way, we can explain why Oscar has beliefs *about* H_2O and why Oscar talks *about* H_2O if he uses the term *water* in Putnam's (1975) twin earth story. His beliefs are about H_2O because his beliefs are sensitive to facts about H_2O, and he talks about H_2O if he uses the word *water* because the content of one's utterances should be explained in terms of the content of one's representations. Oscar's twin is not thinking or talking about H_2O because on twin earth ideal conditions are different; it is not H_2O but XYZ that normally has certain superficial properties, or under ideal conditionals has certain special properties. To account for the intuition that Oscar and not his twin is thinking about H_2O, we have to assume that normality/optimality conditions are determined as is

[17] See also Stalnaker (1984, ch.1)

normal/optimal for *us* in the actual world. We will see later, however, that sometimes the relevant normality/optimality conditions can be set in a different way; it depends on the conversational context how they are determined. According to the above account, the indication relation is context-dependent not only because of the variability of the normality/optimality conditions; determining what the relevant alternatives are also depends very much on the conversational context. We will see later how this double context dependence can be used to account for some puzzling consequences of the information-theoretic approach to content.

Just as the pragmatic analysis of intentionality motivated a coarse-grained account of propositions in terms of sets of possibilities, any externalist strategy working with a notion like *indication* will motivate a rather coarse-grained conception of content, the object of (holistic) attitudes like belief. In particular, it will not allow for a distinction between propositions that have equivalent truth conditions. The reason is that such propositions will behave identically in causal and counterfactual constructions (Stalnaker, 1994). This suggests that the possible world analysis of content is the correct one.

Stalnaker (1984, ch.1) argued that the causal account of intentionality should not replace the pragmatic account, but should only complement it.[18] He argued that the causal backward-looking account should take care of the *content* of beliefs, while the pragmatic forward-looking account should take care of the *functions* that different attitudes have in determining or explaining action.

1.7 Problems for the combination of pragmatic and causal accounts

Stalnaker argued for a combination of the pragmatic and the causal information-theoretic accounts of intentionality. At first blush this seems to be impossible. According to the pragmatic picture we can have false beliefs, and we can have, on a particular occasion, the belief that P, although on this occasion this is not caused by the information that P, but by the information that Q. Doesn't this show that the two accounts are incompatible? I don't think so. The information-theoretic analysis of content sketched above, which appeals to counterfactual dependencies and normal conditions makes no distinction between information and misinformation; rather, the analysis of belief is based on what is *normally* correct. False beliefs are just deviations from the norm. Also, it is crucial that we did not account for content in terms of what was *actually* the cause of a certain information state, but in terms of what *normally* causes this mechanism to be in that state. If normally Q does not cause the representational mechanism to be in a state corresponding with P, the fact that the information that Q actually caused the representational state to be in the particular state that it is in does not demand that the state has content Q.

Still, it seems that the causal account of content leads to unsolvable problems even if the above problems can be accounted for. Once we accept that the content of expressions and intentional states causally depends on external conditions, we are confronted again with many old problems. How can agents seriously believe (doubt) what is expressed by statements whose propositions are necessarily false (true)? We can no longer account for the fact that we appropriately attribute to agents beliefs that seem to be necessarily true or false on the account just defended? Perhaps the most serious problem that it gives rise to is this one: if we accept externalism, then it seems that attitude ascriptions can no longer do the job common sense psychology tells us they do. A commonsense explanation of why the earthling and his counterpart drink so much of the stuff that in their respective communities is called *water* if they are thirsty is that they think that what they call *water* is the best drink for quenching thirst. The problem is that according to the causal conception of content it seems that the belief attribution *Oscar believes that water is the best drink for quenching thirst* is more specific than we want, because we know that Oscar cannot distinguish H_2O from XYZ. Any causal account of content seems to predict a *too specific*

[18] Stampe (1977), Dretske (1981) and Evans (1982) came to basically the same conclusion.

notion of content in these cases. On other occasions the predicted contents seems to be too *unspecific*. Perhaps the most compelling evidence that the externalist position leads to an insufficiently specific notion of content is connected with occurrences of what have been called *essential indexicals*. The following example is from Kaplan (1989) and Perry (1979). Kaplan is looking at a mirror and sees a man whose pants are on fire. This man is actually Kaplan himself, but he does not realise this, and stays cool under the situation. After a while things get hot, however, and he starts to realise that he has been looking at *himself* at the mirror. His earlier coolness disappears, and he shouts *'Help, my pants are on fire!'* How can this change in behaviour be explained if the first person possessive would refer simply to Kaplan?

If content is determined as is predicted via the causal account, it would seem that completely rational agents could have *inconsistent beliefs*. Consider Kripke's (1979) case of Pierre. Pierre grows up monolingually in Paris and learns something from his parents that is expressed by saying "Londres est jolie". On this basis, he is inclined to assent to this sentence. On the basis of the disquotation principle[19] and the assumption that meanings are preserved under translation[20] it seems that we can conclude that *Pierre believes that London is pretty*. Later Pierre goes to England, learns English, settles in an ugly part of London, but he does not realise that the city that he learned about in Paris is the city that he lives in now. He is disposed to utter or assent to "London is not pretty". By the use of the disquotation principle it seems that we may conclude that *Pierre believes that London is not pretty*. On the assumption that Pierre does not give up his earlier belief expressed in French by "Londres est jolie", it is hard to see how we can escape the conclusion that Pierre has inconsistent beliefs if the extension of a proper name exhausts its meaning. This is paradoxical for Pierre may be a perfect logician.

Finally, let us now discuss certain presupposition problems that the causal theory of reference gives rise to. If the communicative actions of rational agents are to be explained in terms of their presuppositions, it seems natural that any agent wants to assert only something informative. An assertion is informative only when the acceptance of the proposition expressed by the relevant utterance eliminates some (but not necessarily all) possibilities representing the speaker's presupposition. Now note that expressions that are interpreted as rigid designators *presuppose* that they have a non-empty extension. Because the extension of such terms exhausts their intension, no proposition can be determined if the expression has no extension. But this gives rise to a new problem: how can we appropriately use statements by which we assert that an expression interpreted as a rigid designator has no (or an empty) extension? Another problem is that it is no longer clear why we sometimes make claims that in our world are necessarily true or false. Normally, a claim makes sense only if it states a contingent proposition. By assuming that for certain expressions the actual extension the expression has exhausts its intension, it also seems to be impossible to explain why certain sentences are always true simply because of the way the words in them are used.

To take an example from Kripke (1972), given that stick S is used to fix the referent of the term *one meter*, we know by definition that the sentence *stick S is one meter long* is true. Still, the sentence is not necessarily true: we can imagine that the stick is longer than it actually is. How can we account for this intuition? Other examples mentioned by Kripke (1971) and Evans (1979) are the use of names like *Jack the Ripper* and *Deep Throat*. If *Deep Throat* is used as a name for the person in the White House, whoever it was, who was the source of Woodward and Bernstein's Watergate information, how can we account for the fact that the clause "*Deep Throat is used as a name for the person in the White House who was the source of Woodward and Bernstein's Watergate information*" cannot

---

[19] If a normal English speaker, on reflection, sincerely assents to 'p', then he believes that p.

[20] If a sentence of one language expresses a truth in that language, then any translation of it into any other language also expresses a truth (in that other language). (Kripke, 1979)

be false? The problem is that although the statement above is, in some sense, necessarily true, the intension of the name depends only on its actual extension which we don't know.

As it happens, the most obvious difficulty we have arrived at will give us an obvious way to resolve at least some of the above puzzles. Although the assumption that the actual extension of certain expressions exhausts their intension can help us to account for the fact that a sentence like *I didn't have to be here, you know* can be true, we can no longer account for the sense in which the statement *I am here, now* is always true.

1.8 Context dependence[21]

What we have missed until now, of course, is the insight that the extension of an expression depends not only on the situations in which its extension is evaluated (the *index*), but also on the *context* in which the expression is used. What is expressed by a sentence is *context-dependent*, so in different contexts the same sentence can express different propositions. Strawson noted this already in his criticism of Russell's description theory, but he concluded that sentences that are context-dependent cannot be handled by formal means. This was overly pessimistic; context dependence, it turns out, can be handled formally if one recognises the importance and distinct roles of context and index. The context partially determines what is said, but does not evaluate whether what is said is true; while the index evaluates only the truth value of what is said. So modal logic should be sensitive to *pairs of situations*, instead of only single situations. The need and possibility of a formal treatment of context dependence by means of a separation of the roles of context and index was recognised by a number of people at about the same time.[22] This liberalisation of modal logic has proven to be increasingly important in the philosophy of language and in natural language semantics. In Kaplan's (1989) theory of context dependence, contexts and indices are entities of different kinds. A context, c, consists of certain aspects of a world, like speaker, hearer, time, etc. For some cases of context dependence, a world also has to be an element of a context. What makes propositions true or false are worlds, and these are accordingly called indices.

Normally, the context in which a sentence is uttered also serves as the index of evaluation. Besides helping to solve some of the puzzles discussed above, there are two reasons why the distinction between context and index is important. The first reason is that in this way we can explain why there are two ways people can disagree about the truth value of a statement. Suppose that the speaker claims something by uttering a sentence, and the hearer disagrees. They can disagree because the hearer has *misunderstood* the speaker. The hearer has made a wrong guess about the context of utterance the speaker was in, and thus about the context-dependent proposition expressed by the speaker. It is also possible that they agree about what is said, but *disagree about the facts* that determine the truth value of what is said. The second reason the distinction between context and index is important is that the distinction makes it possible to handle context-dependent expressions in embedded contexts in a compositional manner without relying on the predicate-logical notion of scope (see Kamp, 1971). Because in normal situations the context of utterance and the point of evaluation of a sentence are the same, it seems that words like *now* and *actually* are superfluous. But they are not, as their occurrence in embedded sentences show. In the following sentences we cannot leave the indexicals out without a change of meaning:

(7) I learned last week that there would *now* be an earthquake
(8) I would like to have more money than I *actually* have.

If two situations are relevant for determining the truth value of a sentence, we might say that the meaning of a sentence is a relation between two situations, a two-dimensional intension. Following Kaplan, we can call this kind of meaning the *character* of a sentence.

---

[21] For a much more extensive discussion, see Zimmermann (1991).

[22] For a short history of the subject, see Van Fraassen (1977).

Stalnaker (1978) calls this relation between two situations a *propositional concept*.[23] For the moment, I will assume that the notions are the same, and I will call this notion a character. The character of a sentence is compositionally determined by the characters of its parts. If *E* is an expression, we might call [E] the character of *E*. Given a context, *c*, [E](c) is the *content* or *intension* of E. [E](c)(i), finally, is the *extension* of *E*, if *i* is an index. The content of a sentence is a proposition, and its extension a truth value. To determine the intensions of (7) and (8) in terms of the intensions of their parts, we have to determine the intension of their embedded sentences with respect to the context of utterance. Double indexing is needed for reasons of compositionality, if words like *now* and *actually* are treated as singular terms or as one-place propositional connectives. In classical one-dimensional modal logic, possible worlds play only one role, so there can be only one kind of rigidity and one kind of necessity. In two-dimensional modal logic,[24] instead, we make a distinction between the role of contexts and the role of indices, so there are *three kinds of rigidity, three kinds of necessity*, and *three kinds of entailment relations*. Let's begin with rigidity. An expression has a *rigid content* if it has a constant content in each context. Indexicals like *now, actually* and *I* are of this kind. If it has the same content in all contexts, it has a *rigid character*. Examples are *king* and *officer*. Finally, an expression is *superrigid* if it has both a rigid character and a rigid content. In English, the logical connectives are examples of this kind of expressions. From now on we will call an expression with a rigid content simply *rigid*, and an expression with a rigid character one with a *constant character*. Now the three kinds of necessity. First, what a sentence expresses in context *c* can be true in every relevant world, [A](c) = K, where K is the set of all relevant worlds. Sentence like *Hesperus is Phosphorus* and *I am Robert* are necessary in this way, because proper names and indexicals have rigid contents. This kind of necessity is sometimes called *metaphysical necessity*. Second, a sentence can be true in every context in which it is expressed. If i(c) gives us the world of *c*, this means that for all *c*: i(c) ∈ [A](c) holds. Some have identified the necessity of such sentences with that of *a priori necessity*. A sentences like *I am here now* is a well-known example of this kind. If a sentence expresses in every context a proposition that is true in every world, the sentence might be called *analytically true*, the third kind of necessity. According to tradition, English sentences like *Tulius is Cicero* and *Every ophalmologist is an oculist* are of this sort.

According to Stalnaker (1978, 1981) we should determine a propositional concept not with respect to sentence types, but with respect to sentence *tokens*. In that case, an utterance context can be thought of as a world containing an utterance token. If I say *You are a fool*, and you misunderstand me, you think that some facts about the world are different than they actually are; I intended to speak about *b*, but you thought I intended to refer to you, *a*, by my use of *you*. If it is assumed that worlds function as indices, and that worlds containing sentence tokens function as contexts, we can liberalise modal logic in the most systematic way and do two-dimensional modal logic. Let us assume with Stalnaker that a world determines both the proposition expressed by a token of a sentence and the truth value of what is expressed. In that case, if *A* is a sentence token, [A] is a relation between worlds. Let's say that w[A]w' means that what is expressed by *A* in *w*, [A](w), is true in *w'*. Let K again be the set of all relevant possible worlds. The proposition expressed by *A* in *w* is, of course,

$$[A](w) \quad = \quad \{w' \in K \mid w[A]w'\}$$

---

[23] If we assume that an expression's character is the same in all possible situations, we might assume that the two notions are closely related to each other. Still, there exists an important difference between the two. A character is a kind of meaning, and associated with a *type* of expression in a certain language. A propositional concept, on the other hand, is not a kind of meaning: propositional concepts are not associated with particular types of sentences of a particular language, but with *tokens* of particular expressions. A Kaplanian context is something like an quadruple containing an agent, a time, a place, and a world, while for Stalnaker a context is a possible world containing an utterance token.

[24] The term is due to Segerberg (1973).

This proposition is known as *the horizontal proposition* expressed by *A* in *w*. For another important proposition, we introduce the diagonal (or *dagger*) operator '†', in the following way:

$$\dagger[A] \quad = \quad \{<w,w'>|\ w'[A]w'\}$$

The *dagger* is a two-dimensional operator which projects the diagonal of the relation [A] into the horizontal:

[A] =	u	v	w
u	1	0	1
v	1	0	1
w	0	1	0

†[A] =	u	v	w
u	1	0	0
v	1	0	0
w	1	0	0

The application of †[A] to any world, determines the proposition that is true in *w'* for any *w'* iff *A* uttered at *w'* is true at *w'*. In other words, †[A](w) is what Stalnaker (1978) calls the *diagonal proposition* expressed by *A* in *w*. Note that for each world *w* and *w'* in K, †[A](w) = †[A](w'). (We will normally forget about the context, and write [†A] as an abbreviation of †[A](w). Because context doesn't play any role, no harm is done.)

With another diagonal operator, ¡, we can express that A is *actually* the case:

$$¡[A] \quad = \quad \{<w,w'>|\ w[A]w\}$$

Note that if w[A]w, ¡[A](w) = K. This operator is normally called the *dthat* or the *upside down dagger*. The *dthat* is a two-dimensional operator that projects the diagonal of the relation [A] into the vertical:

[A] =	u	v	w
u	1	0	1
v	1	0	1
w	0	1	0

¡[A] =	u	v	w
u	1	1	1
v	0	0	0
w	0	0	0

Now we can define three kinds of entailment relations. First, the *classical entailment* relation: *B* follows from *A* in *w*, $A|=_w B$, iff [A](w) ⊆ [B](w).[25] Second, something that might be called *diagonal entailment*: $A|=_d B$ iff [†A] ⊆ [†B]. The strongest notion of entailment might be called *analytic entailment*: $A|=_a B$, iff [A] ⊆ [B]. Note that if $A|=_a B$, then both $A|=_w B$, for all *w* and $A|=_d B$ follow.

Finally, let me note that two immediate generalisations are possible. First, we can define the *dagger* and the *dthat* operators not only on propositional concepts, or characters of sentences, but also on other two-dimensional intensions (see § 1.14). Second, if we use the Kaplanian framework we can relativise the various notions of rigidity, necessity and entailment to particular languages (see Van Fraassen (1979)). Until now we have assumed that statements can be uttered in all contexts. But of course, in every context of use a particular language is spoken. We might assume that in two different contexts the same language is spoken, because in both worlds of the contexts the referents of all (relevant) hidden indexical terms like proper names and natural kind terms are the same in both worlds. So, if we assume that English is spoken, only a limited set of contexts is under consideration. If a more specific dialect of English is under consideration, another set of contexts is assumed. But of course, in one context or possible world, more languages are spoken. However, we can say that one contextual variable says what dialect the speaker is speaking. Thus, language *x* determines a class of contexts *L* if it is assumed that the value of the contextual variable in all contexts in every elements of *L* is *x*. It is now with respect

[25] I am leaving out the quantification over models here.

to *L* that we have to think of the notions of rigidity, necessity and entailment.[26] Finally we can use indexed dthat operators, or backward looking operators.[27] However, in the text I won't make much use of most of these generalisations.

1.9 Solving problems by diagonalisation

We have already seen that distinguishing the roles of contexts and indices makes it possible to explain the distinction in kinds of necessity of true identity statements like *Hesperus is Phosphorus* and *I am here, now*. Let us abbreviate metaphysical necessity by *necessity*, and a priori necessity by *a priori*. If a proposition is not necessarily true or false, it is *contingent*. The first kind of statement is not *a priori*, but if true, necessarily true; while the second kind is *a priori* true, but contingent. So, *I am here, now* is true in every context in which it is uttered, but need not to be true in every index world with respect to a particular context. It follows that *I didn't have to be here, you know* can still be true.

The *a priori* status of the statement *Stick S is one meter long* can be analysed along the same lines. It is useful first to distinguish, with Kripke (1972), two ways an identity statement like *E is the N* can be used. It can be used to state the identity of meaning (intension) of the two terms or to fix the meaning of one term by the meaning of the other. Thus, sometimes the description *the N* in *E is the N* is used to fix the reference of *E*. This is what is going on in a sentence like *One meter is to be the length of S*. The meaning of the name *one meter* is fixed by the reference-fixing use of *the length of S* by the occasion of utterance. In every context in which the meaning of *one meter* is fixed, it will be the length of stick S, although the length of the stick might have been different from what it actually is. Something similar is going on in *Deep throat is the person who was the source of Woodward and Bernstein's Watergate information*. The reference of the name *Deep Throat* is fixed by the fixing-reference use of the description that follows it. The only difference with the foregoing case is that now *we* don't have any specific individual in mind. Whatever the relevant meaning of *Deep Throat* is, it is clear that the counterfactual *If Haldeman had released the information to the reporters, he would have been Deep Throat* is unacceptable because we consider only counterfactual situations in which Deep Throat is the person who *actually* released the information to the reporters (see Evans, 1979).

The next problem we will consider is that of necessary and impossible propositions. How is it possible that a sentence like *I am Robert* can be informative in some conversational contexts? For instance, how can we explain that I can use this sentence as an informative answer to your question *Who are you?* As far as you know, the world might be such that the one you are asking this question is Robert, in *w*, or someone else, in *w'*. Now the above sentence, *A*, uttered by me, Robert, would express a necessarily true proposition, [A](w), in *w*, but it would express a (necessarily) false proposition, [A](w'), in *w'*. According to Gricean conversational rules, and Stalnaker's (1978) first assertion condition, every assertion should express a contingent proposition with respect to what is presupposed by the speaker. The hearer can conclude that the speaker intended to communicate neither [A](w) nor [A](w'). Moreover, on Grice's and Stalnaker's analysis again, it should be clear to the hearer what proposition is expressed by a given sentence. In each possible world of the context compatible with what is presupposed, the same proposition should be expressed. What could this proposition be? In these cases, Stalnaker (1978) suggests, we should look not at the horizontal proposition expressed in a particular context, but rather at the diagonal proposition, [†A]. The diagonal proposition expresses the same proposition in every context because it abstracts away from the context. If [A](w) is necessarily true, but [A](w') necessarily false, [†A] will be contingent. Of course, [†A] can be different from [A] in several worlds only if *A* has a non-constant character. Because the character of *A* is determined compositionally by the characters of its parts, the diagonal

---

[26] Van Fraassen (1979) and Haas Spohn (1994) have made extensive use of this language dependence of characters for the analysis of belief attributions. I won't use it as explicitly as they do.

[27] cf. Saarinen (1978).

proposition expressed by an identity asserted between two expressions treated as rigid designators can be contingent only if these terms do not necessarily have a constant character. This is clearly the case for indexicals; in different contexts it might be a different person who is speaking. In the same way, with a token analysis of diagonalisation, Kaplan's paradox of direct reference can be explained. The problem is to explain how a very slow utterance of *This* [pointing to Venus in the morning sky] *is identical with that* [pointing to Venus in the evening sky] can be informative. This can be explained by saying that in some worlds consistent with what is presupposed in the conversation, the *token* of *this* will not refer to the same object as the *token* of *that*. The result will be that the hearer is informed that the most salient heaveny body in the morning sky is identical with the most salient heaveny body in the evening sky. So, diagonalisation can explain away some paradoxical consequences of the assumption that indexicals and demonstratives are directly referential. Obviously, if it can be assumed that the intension of proper names and natural kind terms are also context-dependent in this way, we could also solve the problem posed by identity statements between two such terms in such a way. But are the intensions of these terms context-dependent in this way?

Of course the intension of a name is context-dependent. A lot of individuals have the same name, and it depends on the conversational context what the most salient individual meant by the use of a name is. What is more interesting to know, though, is whether the meaning of a proper name can also be world-dependent. In one sense it cannot be denied that the meaning of a proper name is world-dependent. It is a contingent fact about our language that Venus was called *Phosphorus*; if the semantic facts about our world were different it might have been the case that, for instance, Mars was called *Hesperus*. But as Frege (1892) stressed, identity statements between proper names need not be about purely semantic facts of our language. So the question is whether we can assume that the meaning of a proper name is world-dependent but not just because of the fact that objects could have been called differently. The externalist theory of reference denies that the meaning of a proper name, or any other kind of expression, is world-dependent in the sense that there is (normally) no description associated with the term that determines its meaning. On the other hand, the causal information-theoretic account suggests that there is a body of information associated with the expressions we use. It can then be assumed that the reference of the expression is world-dependent not because in different worlds it may be a different object or stuff that best fit this body of information, but because in different worlds it might have been a different object or stuff that is the dominant source of this body of information.[28] Remember that according to the information-theoretic account we refer with the English expression *water* to H_2O in this world because we *normally* use this term to refer to stuff that has certain observable properties, and *normally* it is only H_2O that has those properties in the actual world. But we also saw that these normality conditions are contingent; they might be different from world to world. On twin earth the normality conditions of the actual world do not obtain: there it is normally not H_2O but XYZ that has the relevant observable properties, and is 'responsible' for the use of the term *water* by twin-earth (twin-) English speakers. Once it is assumed that the referent of, for instance, a proper name is world dependent, it is clear that by diagonalisation we can normally account for the informativity of an identity statement like *N is M*, where *N* and *M* are both names. In a sense, the reason why the meanings of expressions are world-dependent just depends on semantic facts about the words. Still, we can learn something non-linguistic if we are informed that *N is M*, because even if the exact referent of an expression used in a conversation is not clear, we normally *do* have a pretty good idea about what properties the referents of the used terms have. Thus, if we receive the information that the sentence 'Hesperus is Phosphorus' is true, we learn not only some facts about the semantics of English, but also some astronomical facts. We learn that the most salient heavenly body seen in the morning sky is identical with the most salient heavenly body seen in the evening sky, because we already believe and presuppose that we are in a world in which the referents of the relevant expressions have those properties.

---

[28] See Haas Spohn (1994) for more discussion.

It may seem that once we assume that the reference of a proper name is world dependent, we can also immediately account for negative existential statements containing proper names. But things are not that easy. If the reference of a proper name is world-dependent only because the dominant source of the information associated with our use of a proper name is not clear, we still seem to presuppose that a dominant source of this information does exist. But isn't this exactly what we claim not to be the case with negative existential statements? Perhaps negative existential statements should not be seen as assertions, but as presupposition denials instead. But then, denials are normally reactions to earlier utterances in which the opposite is asserted. This leaves us with the equally difficult question of how we can appropriately assert contingent propositions with positive existential sentences. Donnellan (1974) proposed that with a negative existential statement we simply assert that the proper name has no referent. Stalnaker (1978) offers an attractive way to implement this solution: namely in terms of his diagonalisation strategy. From Donnellan's discussion it seems that negative existential statements involve only the *mention*, rather than the use, of proper names. But when proper names occur in negative existential sentences it seems that the hearer has to understand the singular term in the same way as in a normal use of a proper name. If we use the diagonalisation strategy, we don't have to distinguish different uses of proper names to account for Donnellan's proposal. The normal use of a proper name presupposes an existing individual that is the dominant source of the relevant information associated with the name. Sometimes, however, it might be presupposed to be possible that the actual source is just, for instance, a character in a novel, an object to which the existence predicate does not apply. In this case, the diagonal proposition associated with a sentence like *N does not exist* will be contingent, and seems to be the right candidate for that what is expressed by such a sentence.

1.10 Self-locating beliefs

Traditionally, a belief state is represented by a set of possible worlds. According to Perry, however, there is a problem for the traditional analysis which is related to self-locating beliefs. That is, the traditional analysis cannot account for certain kinds of *sameness of the beliefs* that different agents might have. Consider crazy Heimson (Perry, 1977), who thinks that he is David Hume. Alone in his study, he says to himself, *I wrote the Treatise*. Of course, he has not. So, contrary to the case in which Hume was thinking this thought, Heimson is thinking something false. However, it seems that we can explain some of Heimson's and Hume's behaviour in the same way if they both think *I wrote the Treatise*. How can the traditional analysis account both for the difference of belief and for the fact that some of their actions can be explained in a similar way?

Following Kamp (1971) and Kaplan (1989), we can do so by modelling a belief state not by a set of possibilities (indices), but rather by a function from contexts to such a set of indices, a *character*. We can explain some of the actions of both Heimson and Hume in a similar way because they have a belief in common. They both stand in the belief relation to the character expressed by the sentence *I am Hume*. Their beliefs differ, however, because the *propositions* expressed by this sentence if said (or thought) by Heimson and Hume are different. Modelling a belief state by characters can also account for *fine-grained ignorances*, a notion which will be explained presently.

Traditionally it was assumed that possible worlds could be completely determined by an impersonal description or eternal sentence. Two possible worlds are the same, if they are qualitatively the same. However, Lewis (1979a) showed that belief states cannot be represented by sets of possible worlds understood in this way. Such a representation is not fine grained enough. We should distinguish more possibilities than there are qualitatively different possible worlds:

Consider the case of the two gods. They inhabit a certain possible world, and they know exactly which world it is. Therefore they know every proposition that is true at their world. Insofar as knowledge is a

propositional attitude, they are omniscient. Still I can imagine them to suffer ignorance: neither one knows which of the two he is. They are not exactly alike. One lives on top of the tallest mountain and throws down manna; the other lives on top of the coldest mountain and throws down thunderbolts. Neither one knows whether he lives on the tallest mountain or on the coldest mountain; nor whether he throws manna or thunderbolts. (Lewis, 1979a, pp. 520-521)

Even if two individuals know exactly what qualitative world they live in, they still might lack certain pieces of knowledge. I will say that such agents are ignorant of certain fine-grained pieces of information. But how, then, can the ignorance of the two gods be accounted for? If belief states are represented by characters the problems disappears: the two sentences *I am the god on the tallest mountain* and *I am the god on the coldest mountain* don't express the same character.

Perry's (1977) proposal was adopted in cognitive psychology. According to the research strategy in cognitive psychology known as *individualism*, psychological explanations of behaviour should and can be given completely in terms of the internal states of agents. They *should* because what causes the behaviour of the agents are these states. This doesn't mean that these internal states don't have content. They have contents, but (given what has been learned from the twin-earth stories) these contents cannot be the wide contents, the contents of thoughts determined via externalist means. Different believers can believe different proposition by thinking a thought of the same sentence type. Still, two people who are thinking this have something in common. What they have in common, it is proposed, is a function from contexts to propositions - that is, a character. Thus, psychological explanations *can* be given in terms of internal states only, because of the existence of characters. The contents of internal belief states, the *narrow contents*, are modelled by characters; and what the believer believes, if it is embedded in a specific context, the *wide content* of his belief, is just the result of applying the narrow content to the actual context. Let us denote the internal belief states of Oscar and his twin by [O] and [TO], respectively.[29] The (narrow) contents of their internal states are the same, but because they live in different environments, w and w', the intensions of their thoughts are not the same, $[O](w) \neq [TO](w')$.[30]

According to individualists, belief states should be modelled by something like characters, or better, by sets of characters. But this is problematic if we assume that we should represent a belief state by a set of possibilities, as the pragmatic account of intentionality seems to demand. This is given up, however, when belief states are modelled by sets of characters. According to Lewis (1979a) and Stalnaker (1981), we don't have to model belief states by characters to account for the fine-grained ignorances that the two gods have. If we use diagonalisation we can still model a belief state by a set of possibilities.[31] According to Lewis, the gods know what world they live in, but lack knowledge about *who* they are, or *where* they are in a world. He concluded that a belief state can no longer be represented by a set of worlds, a proposition. Analysing self-locating beliefs, according to Lewis, requires a belief state to be represented by a set of agents, a *property*. The believer has a belief about himself, namely that he possesses a property. This property can be that he inhabits a certain world, but it can also be that he is a certain individual, or that he is in a certain position in a world. Lewis can assume that belief states may be represented by sets of individuals because he assumes that individuals can live in only one possible

---

[29] An individualist like Fodor assumes that a belief state should be modelled not by a character, but rather by a set of characters.

[30] This raises the question of when, according to individualists, a belief attribution is true. According to Stalnaker (1981, n. 11), Perry says that for the *truth* of a belief attribution it is just wide content that counts. Only for the *appropriateness* of the belief attribution is it also narrow content that counts. If this is the right interpretation of Perry, it looks very much like the neo-Russellian analysis of belief attributions proposed by Salmon (1986). If Stalnaker's interpretation is correct, Perry gave up this analysis in Crimmins & Perry (1989).

[31] And note that characters are much more fine-grained entities than diagonals.

world. If we don't want commit ourselves to that assumption, we can say that a belief state should be represented by a set of world-agent pairs, or centered worlds. If <w, a> is such a pair representing an element of the belief state of some individual L

ingens, *a* is the individual that in *w* possesses all the properties Lingens ascribes to himself in the actual world. According to Lewis, *de dicto* beliefs and beliefs with essential indexicals are always self attributions or *de se beliefs*. So, as far as Lingens can tell, he might be the individual *a* in *w*. The information that the two gods lack is not what world they live in, but *who* they are. Their belief states can be represented by the following set: {<gt,w>, <gc,w>}, where *gt* is the god on the tallest mountain and *gc* is the god on the coldest mountain. To analyse other cases of essential indexicals in a similar way: the belief state of John, for example, should in general be represented by a set of quadruples of entities, where such a quadruples, <a, t, p, w>, consists not only of the individual *a* John takes himself to be in *w*, but also of *t*, the time he thinks of as 'now' in *w*, and *p*, the place he takes himself to be in *w*. Because many different n-tuples can contain the same possible world, Lewis' representation of belief states seems to be finer-grained than the pure possible-world account allows for.

According to the pragmatic analysis of attitude states, attitude states are holistic in nature. We do not have a belief box, with several belief objects (however they are modelled) in it, and a different desire box, with several desire objects in it. Instead, the attitude state of an agent is modelled by a global belief/desire state, where the belief determines the relevant possibilities and the desire orders these (and other) possibilities with respect to their desirability. If the possibilities needed to model certain beliefs are finer-grained than possible worlds, the question arises whether this fine-grainedness is also needed for the analysis of desire, and thus for the analysis of deliberation. Both questions are affirmatively answered by Lewis. He convincingly argues that for some deliberations it is important that an agent considers more possibilities than the traditional conception of possible worlds would allow for. It can be that the most useful action to undertake if you are at one place is different from the one you would undertake if you were at another.

By means of Lewis's finer-grained representation of belief states, it is also possible to say what is special about self-locating beliefs. Self-locating beliefs are special in that they crucially involve not only the world of a possibility, but also something else. But how can Lewis account for the fact that we can explain some of Heimson's and Hume's behaviour in the same way when they both believe *I wrote the Treatise*? According to Lewis we can explain their behaviour in a similar way, because we can characterise their belief states in a similar way. But *how* can we do so?

The simplest way would be to follow Lewis and say that both would self-ascribe the same property. Equivalently, as shown by Von Stechow (1984), we might also account for sameness of belief in terms of Kaplan's theory of demonstratives.[32] In Kaplan's theory, sentence type *A* is true in context *c* iff c[A]i(c) holds. But then we can associate with sentence type *A* the set of contexts in which it is true. Let's say that [□ A] denotes this set. A context in Kaplan's theory is a set of quadruple like <a,t,p,w>, a possibility of the same kind as the possibilities used to model a belief state by Lewis. A Lewisian belief state can thus be thought of as a set of Kaplanian contexts. According to Lewis's analysis, Hume and Heimson share a belief because both of their belief states are subsets of [□ I wrote the Treatise]. Sameness of belief can be accounted for not only in Kaplan's framework, but also with two-dimensional modal logic. In two-dimensional modal logic, diagonalisation makes sense because contexts and indices are supposed to be of the same kind. Until now we have followed Stalnaker (1978) in assuming that a context can be represented by a world. But to make contexts and indices of the same type, we can also go the other way round and make indices into more fine-grained entities. When *A* is a sentence, we simply assume that context-index pairs (cip's) both determine *what* is expressed by *A*, and

[32] That is, if we assume that the extension of, for instance, proper names is not world-dependent.

determine the *truth value* of what is expressed by *A*. Let's abbreviate such a cip, <c, w>, by *e*. Thus, the character of *A*, [A], is seen as a relation between cip's (Van Fraassen, 1979). The diagonal expressed by *A* in *e* can now be determined just as before, †[A](e) = {e'| e'[A]e'}. Let us now assume that a context is something like a triple, <a,t,p>, and that the index is a world. This means that a cip is really a quadruple. We have seen that we can read Lewis (1979a) as representing a belief state exactly by a set of such quadruples. What Heimson and Hume have in common is that each of their belief states are subsets of the following set of cip's: [† I wrote the Treatise].

It's nice to know that essential indexicals do not force us to give up the traditional view that belief states can be modelled by sets of possibilities, where these possibilities need not be as unspecific as possible worlds. However, traditionally it has been assumed that a believer stands in a relation to some informational content, a proposition, and that this notion is individuated in terms of truth conditions only. There are at least three good reasons, I believe, why the object of belief should be thought of in this way. First, consider sleeping O'Leary (Stalnaker, 1981) locked up in the trunk of his car. He wakes up when the town clock tolls, but isn't sure whether it rings three or four times. "I wonder whether it is now three o'clock," he thinks. At nine o'clock he is rescued from the trunk of his car. This time he asked himself "I still wonder what time it was *then*". What he wonders about at these two times, it seems reasonable to assume, is the same, a proposition; but on Lewis's account of essential indexicals this reasonable assumption cannot be made. Second, it seems natural to assume that the objects of speech acts and the objects of beliefs are of the same kind, and are propositions, rather than properties. As Stalnaker (1981) noticed, only if the objects of speech acts are propositions can we give a straightforward account of the following kind of conversation:

Heimson, not so sure anymore whether he is Hume, wants to ask the almost omniscient god on the tallest mountain who he is. Finally reaching the top of the tallest mountain he says to the god "I'm confused and don't know who I am," and then asks "Can you tell me? Who am I?" "You're Heimson, the crazy student," replies the god somewhat impolitely.

The proposition expressed by the answer given by the god on the tallest mountain is a direct answer to Heimson's question. Third, and most important, informational content should be individuated by truth-conditional content to be able to behave identically in causal and counterfactual constructions. We have seen that according to the information-theoretic account of intentionality the dispositional concept of *belief* is explained in terms of counterfactual dependencies. Consequently, the object of belief should be individuated by truth-conditional content, a set of possible worlds.

How can we account for the extra fine-grainedness needed to account for self-locating beliefs in terms of possible worlds individuated only by truth conditions? Stalnaker (1981) suggests that this can be done if we give up the assumption that possible worlds, and/or the relations between possible worlds, can be characterised by completely qualitative means. What Lewis's example of the two gods shows is that we need to distinguish more possibilities than worlds that we can distinguish by qualitative means. But then, once we assume with Kripke (1972) that individuals can exist in more than one possible world independent of their (non-essential) properties, we already have to assume more possible worlds than we can qualitatively distinguish. In particular, there is a distinction between the actual world where *d* is the god on the tallest mountain and a counterfactual world, qualitatively indiscernible from the actual world, where *d* is the god on the coldest mountain. The ignorance of *d* can be modelled as a doubt whether he is in what *we* would call the actual world or this counterfactual world. Crucial for Stalnaker's analysis is, first, the observation that agents who have beliefs are inhabitants of the actual world; and second, the assumption that the subject of the attitude exists not only in the actual world, but also in all worlds that help to characterise his belief state. Of course, Stalnaker's solution is closely related to Lewis's. Suppose for simplicity that we model a belief state by a set of world-agent pairs. Suppose that *d* is an individual with a doubt. The belief state of *d* can then be modelled in Lewisean way by something like {<w,a>, <w,b>}. The only

qualitative way in which <w,a> differs from <w,b> is that in the first possibility, *a* is the god on the tallest mountain; while in the second, *b* is the god on the coldest mountain. For the Stalnaker solution, suppose that *d* is the only individual that exists in the actual world, *w*, and this counterfactual world, *w'*.[33] The belief state of *d* can be modelled by {w, w'}. The only way in which *w* differs from *w'* is that in *w*, *d* is the god on the tallest mountain, while in *w'*, he is the god on the coldest mountain. Obviously, there is no substantial difference between the two solutions.[34]

But what about the case of indistinguishably identical twins? Aren't we committed, on a reasonable assumption of supervenience, to claim that their belief states should be represented in exactly the same way? If so, this seems to be a real problem for any externalist position. But then, all we have to explain is that under qualitatively identical circumstances the two twins would act in exactly the same way. This explanation is given in terms of an attributed belief/desire system. We have seen earlier that the pragmatic account of intentionality need not completely determine the content of one's thought. The pragmatic account cannot distinguish two belief/desire states that predict, or can explain, the same kind of behaviour in the same way. So, we can substitute one object throughout the whole belief/desire state for another, without predicting any difference in behaviour (Stalnaker, 1984, ch. 1). That in this case we substitute the *agents* of the beliefs for one another doesn't seem to make a crucial difference.[35]

To account for the case of the two gods in terms of possible worlds only, we don't need to rely on diagonalisation; the descriptions *the god on the tallest mountain* and *the god on the coldest mountain* are not considered to be rigid designators. Things are different if both terms that flank the identity sign are thought of in this way. Consider the following example from Perry (1977).

Rudolph Lingens is the amnesiac lost in the Stanford Library. Lingens knows a lot about himself, but unfortunately he doesn't know that *he* is the amnesiac lost in the Stanford Library. That is, before he has found out, he would not assent to the statement *I am the amnesiac lost in the Stanford Library*. But after reading a biography about himself, he believe that Rudolph Lingens is the amnesiac lost in the Stanford Library. Suppose now that the proper name *Rudolph Lingens* and the indexical *I* are interpreted as rigid designators. It follows that *I am Rudolph Lingens* expresses either a necessarily true or a necessarily false proposition. But how then can we explain in terms of possible worlds only that Lingens is wondering whether this sentence is true or not?

Because the sentence contains two rigid designators, it seems that there are two ways to solve the problem: by giving up the rigidity of either the proper name or of the personal

---

[33] Or *d* and *d'*, two 'individuals' related to each other via a primitive counterpart relation.

[34] That is, if it is assumed that individuals can have singular beliefs about themselves only. If this assumption is not made, it is not clear how belief states can explain behaviour if those states are modelled by sets of possible worlds (this difficulty was raised by G. Soldati and J. Stotz (personal communication). Although this assumption is compatible with Stalnaker's first way of describing the puzzle, it is not a very natural assumption. Only after I have formalised counterpart theory in the last part of this chapter I will be able to offer a solution compatible with Stalnaker's claim that to account for self-locating beliefs, we don't have to abandon the assumption that from the believer's point of view, belief states should be modelled by sets of possible worlds.

[35] A bit more formally. Let *w'* be an element of K(d,w), the Stalnakerian representation of the belief state of *d* in *w*. Is there for this world *w'* a unique world-agent pair <v,a> such that <v,a> is an element of a Lewisian representation of the belief state of *d* in *w* ? If qualitative difference is all what counts for the pragmatic analysis of belief, I think the answer is *Yes*. *w'* corresponds with <v,a> as an element of the representation of d's belief state if for all qualitative properties *P*, P(d,w') iff P(a,v). As mentioned in a footnote above, Stalnaker's solution as I have represented in the main text can only account for the explanation of behaviour if it is assumed that agents can only believe singular propositions about themselves. I don't find this a satisfying solution, but I don't think Stalnaker (1981) is committed to it. See the last part of this chapter for more on this.

pronoun. If we want to describe the situation by giving up the rigidity of the proper name, we can assume with Stalnaker (1981) that Lingens has known all along who he himself is, the same individual in all possible worlds representing his belief state, but did not know who Rudolph Lingens is. Diagonalisation now has the effect that the rigidity of the name *Rudolph Lingens* is given up. Lingens wonders who is the source of the body of information he associates with the name *Rudolph Lingens*.

Let's be a bit more explicit. There are two relevant situations. First we have the actual world, *w*, where *d*, the actual Lingens and the reader of the biography, is also the subject of the biography, and thus the source of the information he associates with the name *Rudolph Lingens*. Second, we have the counterfactual world *w'*, where *d*, the actual Lingens and the reader of the biography, is neither the subject of the biography nor the source of the information he associates with the name. His belief state before he learns that he is Lingens can be characterised by {w, w'}; after he learned this, however, world *w'* is eliminated. Kaplan's (1989) change from cool to hot can be explained in the same way.

I have described the situation in which Lingens is as if he is not wondering who he himself is, he is just wondering who Lingens is. Is it also possible to describe the situation in such a way that Lingens is wondering who he himself is without giving up the assumption that belief states should be modelled by sets of possible worlds individuated by truth conditions only?

Stalnaker (1981) argued that this can be done, too, but only if a token-reflexive analysis of indexicals is assumed. According to a token analysis, a context is not an agent, time, place, and world tuple as in Kaplan (1989), and also not simply a world, but a world plus a token of an expression. The referent is then the referent of the token of the expression in that world. In particular, the referent of a token of *I* in a world is the utterer of this token in this world. However, in different worlds it might have been a different person who was the utterer of the token (or a counterpart of it). Thus, in different worlds the personal pronoun *I* might have been a different referent. Now consider Lingens again; who says to himself *I am Lingens*. According to the first way of describing the situation, Lingens has known all along who he himself is, but hasn't known who the referent of the name *Lingens* is. According to the second way, it is the name *Lingens* that refers to the same individual in all worlds characterising Lingens' belief state, but Lingens is not sure who he himself is, that is, who the utterer of the personal pronoun *I* is. It is possible to describe the situation in this second way, because it can be assumed that when somebody is thinking a thought, the least that the agent believes is that this thought exists. That is, a token of this thought (or a counterpart of it) will exist in all worlds characterising his belief state. But this is enough to let diagonalisation do its work. The result will be that Lingens will believe the proposition that can be expressed by *the thinker of the thought token of 'I am Lingens' is Lingens*.

Stalnaker has given two kinds of solutions to the problem of how to analyse indexical belief on the assumption that belief states are to be modelled by sets of possible worlds. It is important to see that the two proposed solutions do not correspond to different kinds of situations, but rather are two different ways of describing the single situation of Lingens thinking *I am Lingens*. Moreover, Stalnaker claims that the two ways of describing the situation are *equivalent* in the following way: in whatever way we describe the situation, the *diagonal proposition* determined by the token of the sentence *I am Lingens* will be the same.

1.11 Belief and belief attribution

Traditional wisdom has it that the truth value of an attitude attribution does not depend on the *extension* of the embedded sentence. However, given the assumption that a semantic theory maps surface structures of natural language to semantic values in a rigid way, we seem to be forced to accept that the truth value cannot even depend on the *intension* of the

embedded sentence, since this would lead to inconsistent beliefs would not be able to explain agents' behaviour.

Consider now a situation where *we* know how to distinguish H_2O from XYZ, but Oscar does not. Although the belief attribution *Oscar believes that water is the best drink for quenching thirst* can be used to explain Oscar's H_2O drinking behaviour, Oscar himself would not be able to make a distinction between the actual world, where what is called *water* by ordinary English speakers is H_2O, and the counterfactual twin earth, where XYZ is denoted by *water*. In this sense, his thoughts do not seem to be *about* H_2O, although the causal account seems to predict that they are. We have seen that individualists have concluded that what explains behaviour does not depend on something outside of the agent. What explains Oscar's behaviour is a thought internal to Oscar; and what the content of this thought is, the *narrow content*, can be determined without looking at external circumstances.[36] It is then assumed that, by something like the diagonalisation strategy, we can determine what this narrow content is. What we have to ask is what Oscar's thought token *Water is the best drink for quenching thirst* would express according to the semantic rules of Oscar's *language of thought* in different possible worlds. The semantic rules of the language of thought assign to all types of expressions functions from worlds to intentions, Kaplanian characters. According to the semantic rules of Oscar's language of thought, the thought token *water* denotes H_2O in the actual world and XYZ in a counterfactual twin-world. If in both of these worlds what is denoted by *water* is the best drink for quenching thirst, in both of these worlds the belief attribution would be true, and that is why his thought is not about H_2O. But there are two problems with this argument. First, it is not at all clear that narrow content *can* be determined without looking at external circumstances. Second, it is not obvious that we *need* to explain the behaviour of Oscar by abstracting away from external circumstances. The assumption that we can determine narrow content without looking at external circumstances is based on (i) the Fodorian assumption that we can single out thought tokens, and that the types corresponding to these tokens belong to a language of thought which has a particular semantics; and (ii) the assumption that we can determine the specific function from contexts to intensions for each expression of the language of thought without looking 'outside the head'. Yet it is not clear why the first assumption should be true; and, as Stalnaker (1989) stressed, Fodor's (1987) claim that two expressions of the language in thought have the same character if they determine the same truth conditions in every context might be true, but says nothing about how to determine the specific function associated with an expression of the language of thought. So, even if there is something like a language of thought, it is not at all clear how the narrow content of expressions of this language can be determined by looking only at the internal state of the agent.[37]

Also, it is not obvious that we *have* to explain the behaviour of Oscar by abstracting away from external conditions. The problem is that although Oscar's thoughts are, from an externalist point of view, about H_2O, he is not able to distinguish earth from twin earth, or H_2O from XYZ. How can we account for both intuitions if content is determined by causal means? To account for the second intuition, it seems that we need to use the diagonalisation strategy. But don't we in that way predict that the content of one's thought is independent of external conditions? No we do not! According to the causal information-theoretic account of intentionality, even something like the narrow content of one's belief state, the content of the belief from the believer's point of view, is dependent on facts of the environment. But this does not mean that the diagonalisation strategy cannot be used. Remember that according to the causal information-theoretic account, someone believes that *P* means that he is in a certain state that *under normal conditions* he would be in only if *P*. But we have seen that both conditions external to the agent and conditions related to the

---

[36] See Fodor (1987). It should be noted that Fodor defends this position only in this article.

[37] For more on this, see Stalnaker (1989, 1990b). Note also that (i) we can make proper names and natural kind terms context-dependent only if we think of language as an element of the context, and (ii) that the pragmatic, or functional, account of intentionality is essentially externalistic.

internal functioning of the representational mechanism belong to these normal or ideal conditions. There are at least two possible ways to determine these normality conditions.[38] Both require that the external circumstances should be normal, but they differ with respect to the normal conditions related to the internal functioning of the representational mechanism of the agent. The relevant normality condition in this case is the ability *of the agent* to distinguish H_2O from other relevant liquids. According to one way of determining these normality conditions, we look at how *we* are able to distinguish H_2O from those other liquids. In this way of determining content, Oscar's belief is *about* H_2O. But given that we know that Oscar cannot distinguish H_2O from XYZ, this is not the way to determine the content of Oscar's belief from his own point of view. To account for the content of Oscar's belief from his own point of view, we should not determine the relevant normality conditions for the agent according to what is normal for *us*. Instead, the relevant normality conditions should be based on the agent, Oscar's normal abilities to distinguish H_2O from other relevant liquids. If in determining normality conditions we demand that the facts about the agent be *normal for him*, we are, as it were, evaluating the belief attribution from the agent's point of view. It is in these cases that we can use the diagonal proposition expressed by *Water is the best drink for quenching thirst*. The diagonal of the relevant propositional concept will be true in the actual world, where *water* denotes H_2O, but also in a counterfactual world where *water* denotes XYZ, a natural kind that looks exactly like the stuff *we* call *water*. It is this diagonal proposition that seems like a reasonable candidate for representing the psychological content of Oscar's thought.

But if we must determine the normal conditions with respect to Oscar's internal functioning of the representational mechanism, as is normal for him, in order to correctly characterise his belief state, aren't we committed to the claim that Oscar's beliefs are not about H_2O at all? How can we account for the intuition that the belief attribution *Oscar believes that water is the best drink for quenching thirst* is both true and still *about* H_2O? To account for this intuition, we have to remember that according to the information-theoretic account of content, the indication relation is analysed in terms of nested counterfactual conditionals: *if* conditions are normal, and *if* various alternatives to *P* were true, then the believer would be in various alternative states. This analysis suggests that it is not only the relevant normality conditions, but also the set of relevant alternatives that are context dependent. According to Dretske (1970, 1981), knowledge and belief attributions are essentially contrastive. The belief attribution *Oscar believes that water is the best drink for quenching thirst* is true if Oscar is able to distinguish those alternatives consistent with the relevant normality conditions where water is the best drink for quenching thirst from those where it is not, and count only the former as true. The set of relevant alternatives depends on what *we* consider to be normal. In normal situations, only these possibilities are consistent with the relevant normal conditions in which what *we* presuppose about the denotation of *water* holds. Because *we* presuppose that water is H_2O, there will be no relevant alternative considered where water is XYZ; thus Oscar's belief can be said to be *about* H_2O, although he cannot distinguish H_2O form XYZ. Alternative worlds where *water* denotes not H_2O, but XYZ instead, are considered only when critical questions about the theory of meaning are considered.

That the set of relevant alternatives depends on conversational context is also relevant to the analysis of *knowledge* (Dretske, 1970, 1981) and of certain cases in which we can attribute to different agents the *same belief* (Stalnaker, 1984). With respect to the first issue, we must be able to account for the intuition that some knowledge attributions are true, but we must also be able to address certain sceptical doubts. That can be done as follows: A knowledge attribution can be true, because we normally consider only possibilities consistent with what *we* presuppose to be normal. Sometimes, however, one of these

---

[38] Better, there are at least two *kinds* of ways. For each kind of way that I will discuss in the main text, there are still lots of ways of determining the exact normality conditions (see Dretske (1981) and Stalnaker (1984) for discussion).

presuppositions, or channel conditions, is called into question. Once this is done - and this is typically done by a sceptic - more possibilities will become relevant. In such cases we ask more of the agent by presupposing less. The agent must have finer-grained discriminating capacities for the knowledge attribution to be true than he has needed before the relevant normality condition is called into question. To account for sameness of belief, it is crucial, I think, that the relevant alternative possibilities consistent with the normal conditions are only as fine-grained as the conversational context asks for. If in discussing what the agent believes only a few issues are relevant, we don't have to distinguish a lot of alternative states of the world. Suppose that the issue of a discussion is whether two individuals have a belief in common. Suppose also that one agent, *a*, has a more complex representational mechanism than the other agent, *b* has. Agent *a* tends to be in different internal states when *P* is the case and when *Q* is the case, while agent *b* does not. Suppose now that agent *a* is in a state that carries the information that *P*, and agent *b* in a state that carries the information that *P* or *Q*. If we assume that containing the same information is a necessary condition for having the same beliefs, then it seems that there is a difference in belief. Still, in a context where the difference between *P* and *Q* is not relevant, we might say that *a* and *b* have the same beliefs. What we do in those situations is to make the set of relevant possibilities by which we have represented the belief state of *a* as coarse-grained as the set of possibilities used to characterise *b*'s belief state. If we do so, we can say that both beliefs are identical. In this way we can sometimes appropriately say that a human being and a dog have the same beliefs, though different discriminating capacities; and we can also explain why we can sometimes truthfully attribute the same beliefs to two individuals about a certain topic, although one individual is an expert in the field and the other is not.

In the case of Putnam's twin-earth example, the question was how we could account for the intuition that the belief attribution *Oscar believes that water is the best drink for quenching thirst* is both true and still *about* H_2O. To account for the intuition that his belief is *about* H_2O, we needed to determine the relevant normality conditions that are normal for *us*. On the other hand, to make the truth of the belief attribution compatible with the fact that Oscar cannot distinguish H_2O from XYZ, we had to assume that the only worlds that are relevant to the conversational context are those in which it is H_2O that has the superficial properties that H_2O has. In the twin-earth example invented by Burge (1979), we have to follow a different strategy to account for the truth of certain belief attributions. Consider Bert, an English speaker who has arthritis. Unfortunately, Bert does not know that arthritis is, by definition, a disease of the joints. He says, and apparently believes, that he has recently developed arthritis in his thigh, which is impossible. Still, when we know that arthritis is a disease of the joints only, the belief attribution *Bert believes that he has arthritis in his thigh* seems to express a contingently true proposition. How can we account for this from an externalist point of view? Bert doesn't believe that he has what is, by definition, a disease of the joints in his thigh. Also, his belief is intuitively not *about* arthritis. But if his belief is not about arthritis, how can the facts about Bert be consistent with an externalist account of content? But this is not a real problem. True, his thoughts are not about arthritis, but that doesn't mean that his thoughts do not depend on external conditions. His thoughts are about a more general disease, a disease one can also have in one's thigh. Nice, but how can an externalist account for the truth of the above belief attribution? It is here that we also need to make use of diagonalisation in a belief *attribution*. Setting the normality conditions as normal for us will make the belief attribution trivially false, so by Gricean reasoning we conclude that this is not we way we should proceed. We saw above that by using diagonalisation, we set the normal conditions related to the internal functioning of the agent's representational mechanism. In this case there is something wrong with Bert's use of English. In the dialect of English that he speaks, *arthritis* does not denote a disease of the joints only, but a more general disease. So, if we attribute to Bert the belief that he has arthritis in his thigh, we must not determine a propositional concept with respect to the worlds in which the normal conditions with respect to English in our world hold, but with respect to a slightly bigger set of worlds, in which the normal

conditions with respect to the dialect of English that Bert seems to speak hold.[39] For the belief attribution to be true, it must be the case that in the actual world the set of worlds consistent with what Bert believes is a subset of the diagonal proposition of the propositional concept determined above. Just as in the case for Oscar, it also holds here that more belief attributions can be true if we presuppose less about the relevant normal conditions.

The above use of the diagonalisation strategy can also be used for belief attributions involving proper names (Stalnaker, 1987a). Suppose that *N* and *M* are two proper names that in the actual world refer to different objects. Suppose also that *we* associate with the name *M* the body of information *D*, and that John also associates this with this name. Suppose now that I say *John believes that N is M*. Obviously, I don't want to attribute to John the belief in an impossible proposition, so you conclude that I intended to say that John believes the diagonal proposition determined by *N is M*. But then, how do we determine this diagonal proposition? What is expressed by a sentence is context-dependent, and the context of interpretation for the embedded sentence of a belief attribution is, according to Stalnaker (1988), the set of worlds that are consistent with everything presupposed to be believed by the agent.[40] To determine the propositional concept expressed with respect to this context of interpretation, we have to ask, for each of those worlds, what would be asserted by *N is M* if it were uttered in this world. Let us assume that *N* is *Mars*, *M* is *Hesperus*, and *D* is the information that corresponds to the way *we* and John are acquainted with Venus as seen in the evening sky. In some worlds compatible with what we presuppose that John believes, it is not Venus but Mars that is the source of this information. In this case, the diagonal proposition expressed by the embedded sentence in its context of interpretation will be contingent. The belief attribution was appropriate, because this diagonal proposition is true in some but not all of the worlds in the context of interpretation for the embedded sentence. The belief attribution itself is true in those worlds in which the set of worlds that characterise the belief state of John in that world is a subset of the relevant diagonal proposition expressed by *N is M*. If the belief attribution is true, John believes that Mars is the most salient heavenly body seen in the evening sky.

It is sometimes assumed that the diagonalisation strategy can only account for belief attributions where the subject matter of the belief attributed is linguistic in kind. True, the diagonalisation strategy can only account for belief attributions where what is at issue is the relation between a certain representation and its content. But then, not all representations are linguistic representations, thought tokens are representations too. For that reason it doesn't matter whether or not the agent the attributions is about speaks the same language as I, the attributer, do. In the example discussed in the last section, for instance, all that counts is that it is presupposed that John is *acquainted* with Venus in the same way as *we* are acquainted with the source of the body of information that *we* associate with the term *Hesperus*. In this way we can account for the intuition that we have attributed to John a belief in an astronomical fact.

Of course, the diagonalisation strategy might also be used in case the attribution is purely linguistic. To determine the propositional concept expressed by the embedded sentence of the belief attribution *John believes that a fortnight is a period of ten days*, I am not determining for each world in the relevant context what the source of the information is that we associate with the term *fortnight*. The only thing that seems to count in these cases is the description of *what is called 'a fortnight' by ordinary English speakers*. So, the belief attribution is true iff John believes that what ordinary English speakers call *a fortnight* is a

---

[39] See also Van Fraassen (1979) and Haas-Spohn (1994).

[40] What is expressed by a sentence is a proposition, a function from possible worlds to truth values. But the domain of this function is context-dependent. If C represents that what is consistent with what *we* presuppose, and K(o,w) denotes the set of possible worlds consistent with what Oscar believes in *w*, the relevant context of interpretation for the embedded sentence of a belief attribution, and thus the domain of the function expressed by the sentence, (normally) is $\cup\{K(o,w) | w \in C\}$. On this, see Stalnaker (1988).

period of ten days. We can conclude that we can attribute a belief about linguistic practice to agents without explicitly using metalinguistic terms. Of course, if the description that counts is a description about the use of a term in a certain language, it is to be expected that we cannot always translate the sentence by which the belief attribution is made into another language, and expect that with this other sentence we could attribute the same belief to the agent as with the original one. For instance, as noted by Church (1954), we cannot attribute the same belief by the translation of the above sentence in German as with the original sentence. Because the translation of *a fortnight* in German is the same as the translation of *a period of fourteen days*, nobody would (*de dicto*) believe what would be attributed by the translation of the original sentence into German. We can conclude that what is attributed in a belief attribution might crucially depend on the language used in the attribution. The reason is that although beliefs are always about content, sometimes this content might be about form.

Stalnaker (1972, 1984, 1990a) suggested that the problem of mathematical belief can also be partly solved by the diagonalisation strategy. Possible world semantics is committed to the view that there are only two mathematical propositions, a necessarily true and a necessarily false one. But in this way there can be no doubt or error about mathematics; we can doubt only contingent propositions. Some have concluded that this problem shows that belief states cannot be modelled by sets of possible worlds: we must take into account not only *what* is believed, but also *how* the agent represents what he believes. A belief attribution is true if both the content and the form of the embedded clause match the belief state of the agent. But this seems to be the wrong way to think of things; beliefs and doubts are always *about* something, and it is only content that counts. What can this doubt or false belief be *about*? The first thing to note is that fine-grained distinctions between logical truths only make a difference to language using intentional systems. If what is expressed by P and by Q is necessarily equivalent, although the agent believes the one but not the other, it seems that the agent doesn't have the information necessary to see that what is expressed by those two clauses is equivalent. What this suggests is that his beliefs and doubts are not about what would be expressed by P and Q in w, the actual world, but about the *semantic information* necessary to determine what they express. Let A be a logical statement the truth of which is agent x doubts. Stalnaker suggested that x's doubt is not related to the *proposition* that the logical statement actually expresses, but about the relation between *statements* and what they express, the semantic information. If you are in doubt about a mathematical statement, you doubt whether the statement expresses the necessary proposition. The diagonal proposition mirrors this, because in other worlds the semantic rules might have been different. An agent can be in doubt about a mathematical statement if in one of the worlds representing his belief state, the words used in the statement mean something different from what they actually mean. Although it doesn't seem unreasonable to assume that mathematics is about semantic structure, surely mathematics cannot be just about the specific ways in which mathematical statements are expressed. If Ralph and Pierre say to themselves respectively *Seven plus five equals twelve* and *Sept plus cinq fait douze* they intuitively have the same mathematical belief. What can this object of belief be, if it is not a necessarily true proposition? Stalnaker (1972, 1990a) suggests that mathematics is not so much about the relation between particular tokens of sentences and the proposition they express as about the relation between more abstract structures that some but not all mathematical statements share and the proposition they express. On a certain level of abstraction, the above English and French sentences share the same structure, and what Ralph and Pierre have in common is that they both believe that sentences that have this structure express the necessary proposition. This suggestion can be analysed in terms of the diagonalisation strategy, because this strategy accounts for beliefs about the relation between certain representations and the contents of these representations, and these representations need not be particular linguistic entities but can be more abstract

representations too.[41] Of course, this suggestion by itself will not solve the problem of equivalence and deduction posed by the possible world framework:

> It will not save us from mathematical omniscience to any interesting degree. Given a formal system, its axiom wffs, and its rules of wff-formation and derivation, the theoremhood or nontheoremhood of given wffs follows logically. Thus if I am logically omniscient, know the axiom sentences and rules of derivation and sentence formation of a given mathematical system, and if I am given a theorem sentence, I will, as soon as I identify the sentences in question, know that *it is a derivable* theorem sentence. (Powers, 1976, p. 100)

But we have seen above that this problem might be partially solved if we assume with Stalnaker (1984) that deduction is the process of integrating different compartments of one's belief state.[42]

1.12 Limitations of diagonalisation

Diagonalisation is sometimes a useful strategy in accounting for belief attributions where the wide content of the embedded sentence seems to result in a too specific or unspecific notion of belief to explain the agent's behaviour appropriately. But the strategy has an obvious limitation. It can be used only for the analysis of belief attributions with singular terms - for instance, when it is presupposed that the agent believes that a term has a dominant source, or that the information associated with the term by normal members of the linguistic community has a unique most dominant source in the worlds compatible with what the agent believes. Thus, in each of the worlds compatible with what *we* presuppose he believes we should be able to find a unique object associated with the term. As we will see, this is the reason why not all cases of belief attribution where wide content is not specific enough can be accounted for by diagonalisation.

Consider Kripke's case of Pierre again. The problem was that the names *Londres* in French and *London* in English seems to have the same meaning - not only the same extension, but also the same intension. But then, how can we escape the conclusion that Pierre has inconsistent beliefs if he is inclined to say both *Londres est jolie* and *London is ugly*? According to the diagonalisation strategy that we have been using, the answer seems straightforward. In the worlds consistent with what Pierre believes as far as we presuppose, the names *Londres* and *London* denote different cities. Moreover, in these worlds the city called *Londres* is beautiful, but the city called *London* is ugly. That's why, according to this strategy, (9a) is true and (10) is false

(9a) Pierre croit que Londres est jolie.
(10) Pierre believes that London is beautiful.

---

[41] Bäuerle & Cresswell (1984) have argued that the diagonalisation strategy cannot solve the problem of mathematical belief in the following way:" it seems very implausible to suppose that when someone mistakenly believes that 14 + 23 = 47 the belief world of that person is a world in which this expression has a different meaning and expresses a truth. For one may well believe *that* without believing that 14 + 23 = 47. If one believes that the sentence "pigs fly" is true because one believes that "pigs" is the word for birds it cannot be concluded that one believes that pigs fly." I agree that in general belief attributions are not about the relation between the sentential token of the embedded sentences and their semantic values. But this doesn't mean that they never are, nor that they are sometimes about the relation between sets of tokens that share a certain structure and their semantic values. Bäuerle & Cresswell are suggesting (following Cresswell & Von Stechow, 1982) that belief attributions like *John believes that 14 + 23 = 47* are about the actual numbers 14, 23 and 47, and not so much about the language. I am not sure what it means to have *de re* beliefs about numbers, but if it means to have beliefs about the structures shared by certain tokens of expressions on a certain level of abstraction, the two solutions come down to the same thing.

[42] A complementary strategy would be to make a difference between tacit and active beliefs, and say that someone actively believes that A if he tacitly believes A and if A is one of the propositions that he is aware of. See § 1.5 for more motivation, and Fagin & Halpern (1988) and Thijsse (1992) for formal accounts.

In this case, it is natural to assume that the intensions associated with the names *Londres* and *London* by normal members of the French and English linguistic communities, respectively, are the same. It follows that the diagonalisation solution to the puzzle assumes that beliefs are at least partly linguistic. From Church's (1954) discussion, we have seen that the diagonalisation strategy can account for linguistic beliefs without being stated in explicit metalinguistic terms. But we also saw that there was something special about such cases; we cannot always translate the sentence by which the attribution is made into another language and make the same attribution with this translated sentence. It depends on the conversational context whether we can make the same attribution with the translated sentence. Such a translation is not allowed if it is known that the agent associates a relevantly different body of information with some of the terms used. It seems that in this way we can account for the fact that given that we know the facts that Kripke has given us, we cannot infer (9b) from (9a)

(9b) Pierre believes that London is beautiful.

in this conversational context, and thus derive a contradiction together with (10).[43] But we should be cautious here. In the example discussed by Church, translation is not allowed because the attributed belief was completely about the language. In Kripke's case of Pierre, however, this is not the case. The two relevant beliefs that Pierre has are both *about* London; he is just acquainted with London in two different ways, and associates with those two acquaintance relations two different names. But if his beliefs can be said to be really *about* London, it cannot be claimed that translation does not preserve truth value. The best that can be said is that (9b) is not a very natural way to state Pierre's belief in the given conversational context.

So far, diagonalisation can at least still be helpful. But now consider the case of Kripke's Peter. Peter has heard of a great musician named Paderewski. So he is inclined to say *Paderewski is a great musician.* We can conclude

(11) Peter believes that Paderewski is a great musician.

In a different conversational context we learn that Peter has heard of a politician with the name *Paderewski.* We know that he thinks that all politician's are bad musicians, and why should this one be an exception? We can conclude

(12) Peter believes that Paderewski is a bad musician.

In fact, however, the politician named *Paderewski* and the famous musician are the same person. Because Peter would not associate a single individual with the name *Paderewski* in the worlds that might be the actual world as far as he knows, it seems that the diagonalisation strategy cannot be used to account for this example.[44] Not all cases where wide content is too coarse-grained can be accounted for by diagonalisation. This point can and has been made for all kinds of terms that are treated by Kripke and others as rigid designators. For demonstratives, consider the Esa Saarinen example:

Esa Saarinen, on a semester's visit to the Philosophy Department of UCLA, has told his wife that he is going for the next two days to San Diego. As a matter of fact he is partaking in a punk show in downtown L.A., to which he has invited several of his philosophical friends. Unexpectedly Esa's wife has come to see the show too. But she doesn't recognise the heavily transformed Esa as her husband, when he appears on stage and continues to think that he is in San Diego. Kaplan, sitting closely behind her in the audience can then whisper to his neighbour, pointing first at her and then at the person on the stage, "She believes that he/that man is in San Diego." (Kamp, lecture notes)

---

[43] Muskens (1989) argues that we should blame the translation principle for this puzzle.

[44] But one might argue that he does associate with particular *tokens* of the name different individuals in his belief worlds, and thus that diagonalisation might still be used.

Analogous to Kripke's cases of Pierre and Peter we can ask: Does or doesn't Esa's wife believe that Esa Saarinen is in San Diego? Just as in the Kripke cases, no simple answer *yes* or *no* seems appropriate.

We have seen that the causal information-theoretic account of content gives rise to two different kinds of problems. In the twin-earth stories the predicted wide content is for some purposes individuated *too specifically*; while for the cases that we have just discussed the predicted wide content is individuated *too unspecifically*. The diagonalisation strategy seems to be very useful for the former cases, and it can also successfully account for some cases of the latter kind, but we saw that it cannot account for all of these cases. But in those cases where diagonalisation doesn't help, how should we analyse belief attributions where the wide content is too coarse-grained? The problem is an old one: it is the problem of *de re* belief attributions.

1.13 *De re* belief attributions

Quine claimed that modal statements cannot be made about particular individuals, because that would give rise to essentialism. But there are really two questions of essentialism and quantified modal logic. First, if 'F' denotes a non-trivial property that in principle more individuals could have, does quantified modal logic as a formal system commit one to the truth of any formulae like $\exists x,y[\Box Fx \wedge \neg Fy]$? Parsons (1969) showed that quantified modal logic is formally not committed to this *property essentialism*; but if we agree with Kripke and Putnam that natural kind terms are rigid, some formulae of the above form will come out true. Second, to make sense of formulae like $\exists x \Box Px$ and $\exists x Bel(a,Px)$ in the first place, it seems that we must assume that if we quantify over real individuals, these individuals have to be identified somehow with individuals existing in other possible worlds. But how can identification across possible worlds be done if we don't accept that this always comes down to maximal similarity, except by making the haecceitistic assumption that this identification relation is a primitive relation? And what else is haecceitism than an assumption of essentialism? Being convinced of Kripke's argumentation, I do believe necessary statements about particular individuals can be made, and that heacceitism is true.[45] However, if Russell's description theory or the diagonalisation strategy of Stalnaker is assumed, this heacceitism doesn't always lead to counterintuitive predictions for non-extensional contexts, as Quine thought it would. According to Kripke and others, the principles (SI) and (UI) should be valid for *modal* contexts. Together with the description theory, this doesn't seem to create so many problems. Problems would arise if the two principles were also assumed for *attitude* contexts. I have suggested above that some of the problems this posed for singular terms and common nouns could be solved by diagonalisation. Whereas Quine thought that *modal* statements about particular objects are not possible, he admitted that *belief attributions* about particular objects can be made. But he also argued that intensional logic cannot analyse such attributions, because this would lead to inconsistencies. And indeed, the assumption that for a belief attribution about a particular individual, it is always this individual that is referred to in all worlds compatible with what the agent believes leads to embarrassing results.

Quine (1980) admitted that the inference from (1a) and (1b) to (1c), repeated below:

(1a) George IV believes that Sir Walter Scott is Sir Walter Scott.
(1b) The author of *Waverly* is Sir Walter Scott.
(1c) George IV believes that the author of *Waverly* is Sir Walter Scott.

can be blocked by the description theory, but rightly remarked that

---

[45] This doesn't mean that I believe in property essentialism. For discussion, see §1.14.

Referential opacity remains to be reckoned with even when descriptions and other singular terms are eliminated altogether. (Quine, 1980, p. 154).

The problem is that even when all singular terms are eliminated, a modal logician still has to make sense of formulae like $\exists x\mathrm{Bel}(\mathrm{John}, Px)$. The description theory only shifts the problem to variables, and the diagonalisation strategy does not do much better.

Consider Quine's Ralph who, one evening, sees a man with a brown hat whose suspicious behaviour leads Ralph to believe that the man is a spy. On another occasion, Ralph sees the same man at the beach, but he does not recognise him as the same man, and the thought that the man he sees at the beach is a spy does not even occur to him. If we would represent relational beliefs by giving the description wide scope, we can represent his beliefs (13a) and (14a) by the formulae (13b) and (14b):

(13a) Ralph believes of the man with the brown hat that he is a spy

(13b) $\exists x[\forall y[\mathrm{Man\text{-}with\text{-}brown\text{-}hat}(y) \rightarrow y = x] \wedge \mathrm{Bel}(\mathrm{Ralph}, \mathrm{Spy}(x))]$

(14a) Ralph doesn't believe of the man he saw at the beach that he is a spy

(14b) $\exists x[\forall y[\mathrm{Man\text{-}seen\text{-}at\text{-}beach}(y) \rightarrow y = x] \wedge \sim\mathrm{Bel}(\mathrm{Ralph}, \mathrm{Spy}(x))]$

But now the story goes on. In fact, the man with the hat who is later seen at the beach happens to be Ortcutt.

(15) $\iota x[\mathrm{Man\text{-}with\text{-}brown\text{-}hat}(x)] = \iota y[\mathrm{Man\text{-}seen\text{-}at\text{-}beach}(y)] = \mathrm{Ortcutt}$

So we seem to be allowed to infer from (13c) from (13a), and (14c) from (14a):

(13c) Ralph believes ór Ortcutt that he is a spy

(14c) Ralph doesn't believe of Ortcutt that he is a spy

Now, does Ralph believe that Ortcutt is a spy or not? Or better, how can we account for the beliefs attributed to Ralph that seem to be about Ortcutt without concluding that his belief state is inconsistent?

The example of Ralph is very similar to Kripke's examples of Pierre. In the latter case it could be argued that diagonalisation can solve the problem. However, it is clear that for Quine's example diagonalisation won't help. Ralph has never heard the name *Ortcutt*, and so he doesn't associate any individual with the name in his belief worlds, and *we* don't associate a *single* body of information with the name *Ortcutt*, which has a single individual as its source in each world compatible with what Ortcutt believes.

A natural reply to Quine's Ortcutt problem for a Russellian would be to demand that a *de re* attribution can be truly made only if the agent *knows* the object *whom* the belief is about. And indeed, this seems to be the solution Russell gave to the problem of *de re* belief. He claimed in *On Denoting* (1905) that *every proposition which we can understand must be composed wholly of constituents with which we are acquainted.* Thus, he demanded that you can have a belief *about* an individual only if you are *acquainted* with that individual. But acquaintance by itself does not solve the Ortcutt problem. It is reasonable to assume that Ralph is acquainted with Ortcutt. The problem is that he is acquainted with Ortcutt in two different ways, and that he doesn't know that a single individual is the source of those two relevant bodies of information. To solve the Ortcutt problem in a purely Russellian framework, we would have to assume that an agent is acquainted with an object only if such cases of mistaken identity are impossible. This seems to be what Russell had in mind; a subject can be acquainted only with objects with which he is in sensory contact. In the strategy proposed by Russell, believers stand in relations with the *content* of the embedded

sentence, that is, propositions. A *de re* belief attribution is true only if the agent stands in the belief relation to a proposition about a particular object. Such propositions about particular objects are known as *Russellian* or *singular propositions*. An agent can grasp such a proposition, only if he is acquainted with the object that the proposition is about, where acquaintance means that mistaken identity is impossible. We might say, then, that a *de re* belief attribution is false if the agent to which the belief is attributed is not acquainted in this strong way with the object that the belief attribution is about.

Note that given Quine's story, this Russellian account will predict that neither (13c) nor (14c) will be true. Both of them will be false because Ralph doesn't know that a single individual, Ortcutt, is the source of the two relevant bodies of information; he knows the identity of neither the man with the brown hat nor the man seen at the beach. According to this picture, *de re* attributions can be truly made only if the possibility of mistaken identity does not exist. But that condition is very hard to satisfy. The suggestion that being in sensory contact with an object makes mistaken identity impossible is simply wrong. Consider the following example from Evans:

> Suppose a person can see two views of what is in fact one very long ship, through two windows in the room in which he is sitting. He may be prepared to accept 'That ship was built in Japan' (pointing through one window), but not prepared to accept 'That ship was built in Japan' (pointing through the other window). Now suppose we try to describe this situation in terms of the ordered-couple conception of Russellian thought. We have a single proposition or thought-content - <the ship in question, the property of having been built in Japan> - to which the subject both has and fails to have the relation corresponding to the notion of belief. Not only does this fail to give any intelligible characterisation of the subject's state of mind; it appears to be actually contradictory. By constructing cases of this kind, it is not difficult to argue, given the assumption that Russellian thoughts must be representable in the ordered-couple way, that there is very little applicability, and perhaps no applicability at all, for the notion of Russellian thoughts outside Russell's own narrow limits. (Evans, 1982, p. 84)

Russellians must conclude that there are almost no true *de re* belief attributions. This conclusion, however, seems to be false. Suppose we tell only one half of the story. One evening, Ralph sees a man with a brown hat who behaves suspiciously and who he thinks is a spy. Ralph has never heard the name *Ortcutt*, but in fact it is the person named *Ortcutt* who is the suspiciously-behaving man with the brown hat whom Ralph has seen earlier. In these circumstances the following belief attributions seem to be appropriate and true:

(16) Ralph believes of the man with the brown hat that he is a spy.
(17) Ralph believes of Ortcutt that he is a spy.

From both ascriptions we can conclude

(18) There is someone of whom Ralph believes that he is a spy.

In this case, even if Ralph has no discriminating knowledge about Ortcutt, *de re* belief attributions seem to be appropriate.

The Ortcutt puzzle about *de re* belief is the result of the assumption that quantifying-into intensional contexts are not constrained in any serious way or in a much too strong way, as proposed by Russell. The puzzle shows that for the possible-world analysis of *de re* belief attributions, it is important to separate the question of which *actual object* the attributed belief is about from the question of what the relevant *representative* of this actual object is in the worlds compatible with what the agent believes.

According to Quine, the difference between relational and notional attitudes is one of *belief attribution*. There is no difference between relational and notional *beliefs*. In the possible world semantics we are working in, this means that in both cases the embedded sentence denotes a proposition, a set of possible worlds; and that with a *de re* belief attribution we try to characterise the belief state of the agent, just like we do with a *de dicto* attribution.

Thus, the question is how we can account for the intuition that it's not only the individual that the belief is about that counts, but also the way in which the agent thinks about the individual in possible world semantics. The most straightforward way to go about this is to assume that at least in case of *de re* belief attributions the quantifiers range not over real objects, but over all individual concepts instead. In that case, two concepts might denote the same individual in the actual world, but different objects in some of the belief worlds of the agent. In this way, two *de re* belief attributions like (13b) and (14b) no longer lead to a contradiction. However, Kaplan (1969) noted that Quine's requirement for 'a conception of the individual the belief attribution is about' would make *de re* belief attributions too easily true. It is counterintuitively predicted that the *de re* belief attribution *Ralph believes of Ortcutt that he is the shortest spy* is true just because Ortcutt actually is the shortest spy and Ralph (*de dicto*) believes that the shortest spy is the shortest spy. Kaplan concluded that it is not enough for a *de re* attribution that the agent has a name or conception that happens to fit the individual whom the belief attribution is about in the actual world. Fit is not enough; the agent has to be *acquainted* with the individual whom the belief attribution is about because of some *causal relation*. In this way the problem of *the shortest spy* is resolved. Kaplan's solution is different from Russell's, because for Kaplan it is possible that a subject can be acquainted with an object in two different ways, such that he doesn't know that a single individual is the source of the two acquaintance relations. In this sense he follows Quine; however, since there is a sense in which Ralph does and one in which he does not believe of Ortcutt that he is a spy, we don't have to conclude that Ralph has inconsistent beliefs.

Kaplan's insights were remarkable. According to him, a belief can be about an individual, in our case Ortcutt, if two conditions are satisfied. First, the agent has a representation of the individual; and second, this representation must be causally connected with the actual individual the belief is about - that is, Ortcutt. By the first condition, we can explain why the agent is disposed to perform certain actions that involve Ortcutt; and by the second condition, we can explain how he came to have beliefs about Ortcutt. But we have seen above that these two conditions not only have to be satisfied in cases of *de re* belief ascriptions. Beliefs are always dependent on the environment, but in the case of *de re* belief, the causal relation, the *acquaintance relation*, is a very specific one. According to the causal-pragmatic account of intentionality, in all cases where a system represents or is about something else, the two conditions for aboutness demanded by Kaplan should be fulfilled. My conclusion is that Kaplan's analysis of *de re* belief attributions is not ad hoc, as is suggested by Schiffer (1990), but part of a very general strategy to explain the notion of intentionality.[46]

We have seen that the diagonalisation strategy seems adequate when the wide content of the embedded clause of the attribution is too specific, but that at least sometimes the strategy doesn't seem to work when the wide content is not specific enough. What I want to suggest now, unsurprisingly, is that in all these cases the *de re* strategy should be used. Note, for instance, that Kripke's (1979) puzzle's about Pierre and Peter are similar in almost every respect to Quine's puzzle about Ralph; and I will assume with Lewis (1981b) that the solution to the latter puzzle can, and probably should, be used to account for the former puzzle. This might be inconsistent with Kripke's claim that the two relevant belief attributions should be analysed *de dicto*, but I don't see why this is so. In the same way, the problem posed by the Saarinen example should, I think, be handled by the *de re* strategy, as indeed is proposed by Von Stechow (1984).

---

[46] But Shiffer's problem with the usual analysis of *de re* belief attribution is not only that it is ad hoc; he also sees a problem with respect to compositionality. But I believe this problem is solved once we admit that the content of the embedded sentence is context-dependent. There need not be a specific content expressed by the embedded sentence; all that is needed for compositionality is that for every context we can determine a specific content expressed by the embedded sentence. See Van Fraassen (1979) for more on this.

We have seen above that Kaplan's analysis of *de re* belief attributions fits the causal pragmatic explanation of intentionality.[47] The content of what somebody believes should be explained in terms of counterfactual dependencies that hold, under certain normal conditions, between the belief state of the agent and his environment. This is the case both when we look at content from the agent's point of view and when the relevant belief is really *about* the actual referent of a term used to characterise the agent's belief. Still, there is an important difference. In case the belief is not really about the actual referent of the term, as with Bert's *arthritis*, the agent's mental state tends to be sensitive not just to the actual referent of the term, but to all that superficially looks like it. In case the agent's belief is really about the actual referent of the term, in our case Ortcutt, the beliefs of the agent according to which we can explain those actions of his that involve Ortcutt are sensitive primarily to facts about the real Ortcutt, and not to individuals that have a lot in common with Ortcutt. How should we account for this *aboutness* in possible world semantics? It is clear that, in the case of Ralph and Ortcutt, there are worlds consistent with what Ralph believes about the actual world in which there are two individuals whom *we* would call Ortcutt if it were not for the other individual. In these worlds there are two individuals who are sensitive to facts about the real Ortcutt. The question 'which individual in this belief world is the real Ortcutt?' has no clear answer. Once we allow that this question need not have an answer once and for all, cross-identification doesn't have to be a matter of strict identity anymore. Individuals of different possible worlds might be identified even if they are not strictly the same, or don't share the same individual essence.

Is this view compatible with the observations and arguments made by Kripke (1971, 1972)?[48] Kripke argued quite convincingly that it doesn't make much sense, if I say 'Suppose Nixon had lost the election', to ask whether a man in a counterfactual world resembles Nixon enough to be his counterpart in this counterfactual world. The argument is quite convincing, but this doesn't mean that the *name* 'Nixon' refers to the actual Nixon in all imaginable possible worlds considered as contexts, nor that *all* possible worlds are being considered when you make the *supposition* that Nixon has a property that he actually does not have.

Normally it is assumed that to determine whether a statement is metaphysically necessary in *w*, we have to look at all possible worlds. The metaphysical accessibility relation for each world gives us the set of all possible worlds; the accessibility relation is universal. Muskens (1989) proposed that we give up this assumption. He assumed that the referent of a proper name is world-dependent, and that in all metaphysically accessible worlds the name refers to the same object as in the actual world. By stipulating that in all metaphysically accessible worlds a proper name refers to the same object, he can account for the observations of Kripke (1972). By giving up the assumption that the set of metaphysically accessible worlds is a superset of an agent's set of doxastically accessible worlds, and by making the referent of a proper name world-dependent, he can account for the fact that an agent can at the same time assent to 'a = a', but deny 'a = b', although *a* and *b* actually refer to the same object.

I think that it is a good idea not to make the metaphysical accessibility relation a universal accessibility relation, and I believe this follows from the information-theoretic account of content, and is compatible with crucial observations made by Kripke (1972). As I argued above, when we check the truth of necessity statements we consider only worlds in which the relevant normality conditions of the actual world hold. Twin-earth stories make it very clear that these normality conditions are contingent, and do not hold in all worlds. How are the normality conditions determined? Following Kripke, we can say that proper names (Nixon), common nouns (cat, water), adjectives (hot, yellow) etc. somehow have actual extension, and that features of the members of this extension determine the features that are

---

[47] Stalnaker (1988) argues that only when Kaplan's analysis is embedded in such a more general account of intentionality, the analysis is not ad hoc.

[48] In fact, even Kripke (1972) suggested that we need some kind of counterpart theory (see pp. 50-53)

essential to the individual, kind, or phenomenon.[49] For necessity statements we consider only counterfactual possibilities where there is an individual, a substance, or a phenomenon that has these essential features. It is irrelevant what these objects or phenomena are called in these counterfactual possibilities.[50]

This suggests that the accessibility relation for determining metaphysical necessity is context-dependent. In particular, it depends on the terms used in the sentence that we want to evaluate what the accessibility relation is. Suppose we want to evaluate whether *Nixon is P* is necessarily true. Then the set of metaphysically accessible worlds consists only of counterfactual worlds in which the person *we* call Nixon exists.

But given our two-dimensional view, something more can be said about the meaning of certain words. According to Kripke, meaning is *intension*, and what fixes the reference is not meaning at all, because meaning has something to do with essential properties. For instance, the description that fixes the reference of *Deep Throat*, the one who released the information to the reporters, is not an essential property of the referent of *Deep Throat*; the referent of the name *Deep Throat* could have done something different. Still, the description has intuitively something to do with the meaning of the word, because the sentence *Deep Throat is the one who released the information to the reporters* is *a priori* true in English. Obviously, the description can be called the *character* of the name - in this case, a function from a context-world to a rigid individual concept. For common nouns and adjectives, things are not really different. The idea is that we associate with a word like *cat* a meaning (character) like *a creature that normally behaves in such and such a way*, which determines its actual extension; and features about the members of the extension determine the features that are essential to the kind (Stalnaker, 1979).[51] These essential properties of cats must be discovered by theoretical and empirical investigation. Thus, if *P* is an essential feature of cats, any actual cat will be necessarily a *P*. However, the essential properties associated with the word *cat* are contingent; in a counterfactual world it might be a different set of creatures that normally behave in such and such a way, and features about the members of this set might determine a set of features that are essential to the kind different from those in the actual world. In two-dimensional modal logic, sentences are interpreted with respect to two possible worlds, the context world and the index world. The meaning of the sentence is compositionally determined by the meanings of its parts. Thus, the parts are also interpreted with respect to two possible worlds. It is sometimes assumed that for primitive predicates only one of those worlds is relevant. From the argument above it follows that we must disagree. Even for a primitive predicate *P*, $I_{w,w'}(P)$ need not be the same as $I_{w',w'}(P)$ or $I_{w,w}(P)$.[52] For instance, let *w* be the actual world, and *w'* be Putnam's twin-earth. Then $I_{w,w'}$(Water) will be the stuff on twin-earth that has the chemical structure of H_2O, and $I_{w',w'}$(Water) will be the stuff on twin-earth that has the chemical structure of XYZ. It should be noted that Kripke has objected to an account that associates with predicates such vague meanings as I have given above:

"'yellow' does not *mean* 'tends to produce such and such a sensation'; if we had different neural structures, if atmospheric conditions had been different, if we had been blind, and so on, then yellow objects would have done no such things." (Kripke, 1972, p. 354)

---

[49] See also Stalnaker (1979).

[50] Van Fraassen (1977) argues that the same holds for physical (logical, etc.) laws. A sentence is physically (logically, etc.) necessarily true, if it is true in all physically (logically, etc.) accessible worlds. A counterfactual world is only physically (logically) accessible if all elements of the suitably chosen (?) set of so-called *law sentences* that hold in the actual world also hold in this counterfactual world. Thus, a sentence is physically (logical, etc.) necessarily true if it follows from the set of law sentences.

[51] To account for this, properties, of course, cannot be defined as functions from possible worlds to sets of individuals. Possible worlds must be defined entities (Stalnaker, 1979).

[52] Why is $I_{w,w'}$(water) not the same as $I_{w,w}$(water)? The reason is that what counts are the essential features of the stuff that in *w*, the actual world, is called water - that is, being H_2O - not the individual molecules that instantiate this stuff in the actual world.

However, Kripke based his criticism on the assumption that *meaning* simply is *intension*. For me, however, meanings are in the first place *characters*. So although superficially the account that I have argued for is in conflict with Kripke's view, it does share Kripke's basic insight that the properties that are essential to, say, cats, are not determined by the writers of the dictionary, but are discovered by empirical investigation.[53]

However we determine the normality conditions relevant to the checking of necessity statements, when we assume that the metaphysical accessibility relation is not the universal relation we need no longer assume that cross-identification is *always* a matter of strict identity. Worlds that help to characterise a belief state need not be stipulated as counterfactual situations in which the referents of referential expressions are the same as in the actual world. The question 'Which individual in a world that Ralph believes might be the actual one corresponding to the actual Ortcutt?' need not have a trivial answer anymore.

If in some of Ralph's belief worlds there are two individuals that are sensitive to facts about Ortcutt, which one do we refer to by a belief attribution like *Ralph believes that Ortcutt is a spy*? That depends, according to Van Fraassen (1979), Von Stechow (1984), Stalnaker (1988), and Crimmins & Perry (1989), among others, not so much on the belief state of the agent itself as on the intention of the speaker and the conversational situation in which the belief attribution is made. Although it seems clear that what is normally *communicated* by a *de re* belief attribution depends partly on the communicative situation, it is questionable whether it also determines the *truth value* of the belief attribution. Especially the example given by Richard (1983) suggests that a *de re* belief attribution is already true if the agent believes the proposition expressed by the embedded sentence under *at least one* representation of the individual the belief attribution is about:

> Consider *A* - a man stipulated to be intelligent, rational, a competent speaker of English, etc. - who both sees a woman, across the street, in a phone booth, and is speaking to a woman through the phone. He does not realize that the woman to whom he is speaking - *B*, to give her a name - is the woman he sees. He perceives her to be in some danger - a runaway steamroller, say, is bearing down upon her phone booth. *A* waves at the woman; he says nothing into the phone. [...] If *A* stopped and quizzed himself concerning what he believes, he might well sincerely utter:
>
> (3) I believe that she is in danger.
>
> but not
>
> (4) I believe that you are in danger.
>
> Many people, I think, suppose that [...] [these sentences] clearly diverge in truth value, (3) being true and (4) being false. [...] But [this] view [...] is, I believe, demonstrably false. In order to simplify the statement of the argument which shows that the truth of (4) follows from the truth of (3), allow me to assume that *A* is the unique man watching *B*. Then we may argue as follows:
>
> Suppose that (3) is true, relative to *A*'s context. Then *B* can truly say that the man watching her - *A*, of course - believes that she is in danger. Thus, if *B* were to utter
>
> (5) The man watching me believes that I am in danger
>
> (even through the telephone) she would speak truly. But if *B*'s utterance of (5) through the telephone, heard by *A*, would be true, then *A* would speak truly, were he to utter, through the phone
>
> (6) The man watching you believes that you are in danger.
>
> Thus, (6) is true, taken relative to *A*'s context.
>
> But of course,
>
> (7) I am the man watching you
>
> is true, relative to *A's* context. But (4) is deducible from (6) and (7). Hence, (4) is true, relative to *A's* context. (Richard, 1983, pp. 439-441)

A more straightforward problem for the view of Crimmins & Perry and others is that it is not clear how they would account for a *de re* belief attribution like *Everybody who has ever seen Ortcutt believes of him that he is a spy*. Although I'm not completely convinced that we shouldn't make the content of the embedded clause of a *de re* belief attribution context dependent in the way suggested by Crimmins & Perry and others, it seems the safest bet to

---

[53] For related views, see Stalnaker (1993), Haas-Spohn (1994, ch. 3), and Spohn (ms.).

adopt the more traditional analysis of Kaplan (1969) to get rid of the context dependence by existential closure.

To account for Quine's puzzle about Ralph in possible world semantics we seem to be forced to assume that one individual in the actual world might have two or more representatives in other counterfactual worlds. This conclusion is also suggested by some other puzzles. Consider Gibbard's (1975) case, for instance, of a statue of Caesar that is made of a lump of clay. It seems that the statue of Caesar could have been made of a different lump of clay, and that the lump of clay that actually constitutes our statue could have been the material ingredient of a statue of Cleopatra. So it appears that there are two distinct 'objects' in a counterfactual world in which both hypothetical situations are realised, that are the same in the actual world.

If the problem of cross-identification is not simply a matter of strict identity between individuals 'existing' in different worlds, it becomes possible that the cross-identification relation is not an equivalence relation. In particular, it need not be transitive. The following puzzle discussed by Chisholm (1967) suggests that indeed the cross-identification relation should not be transitive. It seems natural to assume that after I have exchanged the front-tyre of my bike the result would not be a different bike, but the same bike slightly different. The same holds after I have exchanged its back-wheel, its chain, its bell, its frame, etc.[54] Here is my bike in the actual world, w_0, consisting of n parts, and in w_1 we have a bike consisting of the same parts except that it has a different front-tyre. Intuitively, they are cross-identified. But now we have a whole sequence of worlds $\langle w_0, w_1,...,w_n\rangle$ such there exists a bike in each w_i and w_{i+1} that are made of exactly the same parts, except that one part is different. Intuitively, for each w_i and w_{i+1} it holds that the bike in w_i should be cross-identified with the bike in w_{i+1}. But if the cross-identification relation would be transitive this would have the paradoxical result that the bike in w_0 is also cross-identified with the bike in w_n, although the former bike has no parts in common with the latter bike. Note that those who seek to reduce cross-world identity to genuine identity are in trouble to account for such examples. They have to assume that for one particular *i*, the first *i* bikes are the same, and that the bikes at worlds w_{i+1} until w_n are also the same, but that the bikes at worlds w_i and w_{i+1} are not. The problem is that it goes unexplained why the distinction lies between the bikes at worlds w_i and w_{i+1}, and not somewhere else.[55]

1.14 A Double-Indexing Counterpart semantics for Modal Logic

Now we want to formalise quantified modal logic in such a way that at least most of what has been argued for in this chapter can be accounted for. In the formal semantics that I will give in this section, I will implement the following ideas: (i) it is a semantics in which we can make sense of contingent identity by means of counterparts; (ii) whether two individuals are counterparts of each other does not depend on qualitative features of the individuals, but is a primitive relation and thus compatible with (one version of) heacceitism; (iii) identity statements across possible worlds can be made, but what is expressed by such statements is world-dependent; (iv) although we can make sense of contingent identities, the formula $\forall x,y[x = y \rightarrow \Box(x = y)]$ will still be valid such that '=' stands for the ordinary identity relation that satisfies Leibniz's law; (v) indexicals and proper names will be treated as rigid designators whose interpretation depends only on the relevant reference-context; (vi) although this is accounted for by double indexing, scope differences still play their role; (vii) the informativity of identity statements between two rigid designators, or of a negative existential statement, can be accounted for by diagonalisation; (viii) a distinction is made between metaphysical and epistemic necessity,

---

[54] You don't agree. You say that the frame of a bike identifies it. If so, you should think of another example.

[55] To assume that they are all distinct is counterintuitive, and to assume that they are all the same gives rise to even greater problems. For instance, you have to assume that there are no qualitative constraints at all on cross-identification. (for discussion, see Chisholm, 1967, and Lewis, 1986).

and thus between the horizontal and the diagonal proposition expressed; (ix) the semantics accounts for *de se* attitudes in such a way that what is expressed by a sentence denotes a proposition, a set of possible worlds; and finally (x) it accounts for the intuition that we know *a priori* that the stick that fixes the reference of 'one meter', whichever that is, is one meter long, although it is only a contingent fact that this stick is one meter long, and for the intuition that the objects denoted by common nouns (*light*), adjectives (*yellow*) and natural kind terms (*water*) have essential features (being a stream of photons, having the wavelength of x, being H_2O), although what those essential features are is contingent.

To account for (v) - (x), it is clear that we should make use of double indexing and diagonalisation. To account for this in a Stalnakerian way, both context and index will play a role in the interpretation of primitive terms and predicates. Moreover, for the analysis of indexicals and demonstratives, contexts will be thought of as tokens in a world. We have seen above that we also have to make use of counterparts for the case of demonstratives used in embedded sentences. It is sometimes thought that this rules out a double-indexing account of demonstratives, but I don't see why that should be so.

It is less clear how we should account for cross world 'identity', but let us first remember what such an account should be able to analyse. What we want to account for is that in Quine's story there are worlds consistent with what Ralph believes about the actual world, in which there are two individuals *we* would call Ortcutt, if it were not for the other individual. Also, we want to account for the intuition that although in the possible worlds consistent with what Pierre believes there are two different cities called "London" and "Londres", respectively, Pierre is talking *about the actual London* when he is expressing his beliefs by either *London is ugly* or *Londres est jolie*. These examples show that one actual individual might have two different representatives in a counterfactual world. Similar examples suggest that there might be counterfactual worlds where two actual individuals are represented by just one object. It follows that the cross-world identity relation cannot be reduced to the simplest identity relation that satisfies Leibniz's law; the relation cannot even be an equivalence relation. This doesn't mean that the examples suggest the implausible hypothesis that two different things might be identical in a counterfactual world, nor that individuals are only contingently self-identical. The identity relation between individuals that satisfies Leibniz's law need not be given up. It is only that the question whether two individuals, a and b, in different possible worlds *are* identical need not be the same question as whether these two individuals *represent* the same individual in a certain world.

It is important to realise that when it is assumed that two different individuals, b and c, in a counterfactual world are representatives of a single individual, a, in the actual world, the semantic rules for quantification have to take this into account, too. If a is the only individual in the actual world and is not a P, and this counterfactual world is the only other accessible world, and in this world b but not c satisfies P, we have to decide whether formulae like $\exists x \Diamond Px$ and $\exists x \Box Px$ are true in the actual world or not.

If we want to allow for non-trivial ways in which individuals can be identified across possible worlds, in principle there are, I believe, two ways to go. We have already looked at the first possible strategy; instead of quantifying over real individuals we quantify over *individual concepts*. I have already discussed the proposal for quantifying over all individual concepts, and concluded that we should not do so because it would give rise to implausible consequences when universal instantiation (UI) is assumed. This suggests that we should follow Hintikka (1969) in restricting the set of individual concepts (or individuating functions) over which we quantify. The domain of quantification still contains non-rigid individual concepts, but it doesn't contain concepts that are associated with definite descriptions like *the winner* or *the president of the USA*. Note that once the domain of quantification contains non-rigid individual concepts the formula $\forall x,y[x = y \rightarrow \Box x = y]$ will not be valid anymore; (SI) is given up. This is not as serious as it seems, because the formula doesn't say what it seems to say. In particular, the formula doesn't say

that actual individuals satisfy Leibniz's law. For that reason, its negation, $\exists x,y[x = y \wedge \Diamond x \neq y]$, is satisfiable. The reason is that existential formulae with modal operators inside the scope of the existential quantifier are very easily made true. *Too* easily true, I believe. Also, to account for the intuition that *Ralph believes that Ortcutt is a spy* and *Ralph doesn't believe that Ortcutt is a spy* are incompatible with each other, we have to represent them by $\exists x[x = \text{Ortcutt} \wedge \text{Bel}(r, \text{spy}(x))]$ and $\neg\exists x[x = \text{Ortcutt} \wedge \text{Bel}(r, \text{spy}(x))]$, respectively. I find the representation of the second sentence somewhat unnatural, and prefer a formulation where the negation is put directly in front of the *believe* predicate. Also, if it is assumed that we normally quantify over concepts, the domain of quantification has to be very context dependent. To me it is not clear how to build this context-dependence into the semantics. Moreover, it is not clear how Hintikka could account for a *de re* belief attribution like *Everybody who has ever seen Ortcutt believes of him that he is a spy*. Mainly for the last reasons I will not take over Hintikka's analysis.

The second way to go if we want to identify individuals across possible worlds in a non-trivial way is to follow Lewis (1968) and use a *counterpart theory*. According to Lewis, individual *a* in world *w'* is a counterpart of individual *b* in world *w* if *a* is one of the individuals in *w'* that has most of the properties in common with *b* in *w*. Lewis's counterpart theory based on *similarity* is motivated by his metaphysical doctrines of *possibilism* and *anti-haecceitism*. Possibilism suggests (but does not entail) that individuals exist in only one possible world. According to anti-haecceitism individuals are to be completely thought of as bundles of qualitative properties, and have no underlying *thisness*. Qualitative properties are all we have, and only with the help of these properties can we determine how individuals across possible worlds are to be identified. Of course, for those like me who believe in *actualism* and *haecceitism* there is no good reason to follow Lewis. The notion of similarity doesn't seem to play as important a role in identifying individuals across worlds as Lewis suggests. Consider the following example of Plantinga (1974):

(19) It is possible that in five years my brother will resemble me (as I am now) more than I will.

Although my brother will be more similar to me now than I will be then, it's counterintuitive to assume that my brother then will be a counterpart of me now.[56] Plantinga, following a remark of Kripke's (1972, n. 13), argued that the use of counterparts is in conflict with the way we use modal concepts in natural language. Like Hazen (1979), I don't believe that the second reason for rejecting a counterpart theory is a serious one. I think it is inevitable that we must identify individuals across possible worlds in a non-trivial way, as even Kripke (1972, pp. 271-273) agrees:

> Does the 'problem' of 'transworld identification' make any sense? Is it *simply* a pseudo problem? The following, it seems to me can be said for it. [...] given certain counterfactual vicissitudes in the history of the molecules of a table, *T*, one may ask whether *T* would exist, in that situation, or whether a certain bunch of molecules, which in that situation would constitute a table, constitute the very same table *T*. If statements about [tables] are not *reducible* to those about other more 'basic' constituents, if there is some 'open texture' in the relationship between them, we can hardly expect to give hard and fast identity criteria.

However, this doesn't mean that the relevant counterpart relation should be explained in terms of qualitative similarities and differences. It is this proposal of Lewis's (1968) that Kripke (see pp. 271-273) attacks:

> So: the question of transworld identification makes *some* sense, in terms of asking about the identity of an object *via* questions about its component parts. But these parts are not qualities, and it is not an object resembling the given one which is in question.

---

[56] Lewis (1986) responded by saying that the notion of similarity is context-dependent.

XYZ is not the counterpart of our H_2O on twin-earth, although it has superficially the same properties and is used in the same way on twin-earth as H_2O is used on earth. I agree with Kripke, and will assume with Stalnaker (1987b) that counterpart relations are not defined in terms of a notion of similarity, they are just *primitive relations.*

Not only should we reject the conceptual background of Lewis's counterpart theory, we should also reject his particular way of formalising counterpart theory. On Lewis's formalisation, a formula like $\Box Rab$ would be true in a world if in every world containing counterparts of *a* and *b*, every counterpart of *a* bears the relation *R* to every counterpart of *b*. But, as noted by Hazen (1979), this gives rise to three closely related problems. First, it allows both $\Box Rab$ and $\sim\Box\exists xRax$ to be true in the same world, because there might be metaphysically possible worlds in which the actual referent of *b* has no counterpart. Second, neither $\forall x,y[x = y \rightarrow (\Diamond x \neq y \leftrightarrow \Diamond y \neq y)]$ nor $\forall x,y[x = y \rightarrow \Box x = y]$ is predicted to be valid.[57] Third, Hazen argues that the death of Caesar could, according to Lewis's theory, not be *essentially* of Caesar:

> Suppose in some possible worlds there were two counterparts of Caesar, living in opposite hemispheres of the globe. Each might be related appropriately - by dying it - to some counterpart of the death of Caesar, but neither could be related appropriately to the other's death. Thus neither counterpart of the death of Caesar is *of* all the counterparts of Caesar; so, if Lewis were right, the death of Caesar could not be *essentially* of Caesar. (Hazen, 1979, p. 329)

With Hazen I conclude that we should count a formula like $\Box Rab$ only true in a world, *not* if in every world containing counterparts of *a* and *b*, every counterpart of *a* bears the relation *R* to every counterpart of *b*, *but* if for every relevant way of picking out counterparts, it holds that for every world in which *a* and *b* both have a counterpart, the counterpart of *a* bears the relation *R* to the counterpart of *b*. In other words, we should quantify not over counterparts, but rather over ways of picking out counterparts. Moreover, each way of picking out counterparts will really be a *function* from individuals and worlds to individuals (or representatives) in that world.

In the formalisation I will follow the Fregean assumption that all definite noun phrases should be treated as singular terms, and the suggestion of Stalnaker (1988) that the proposition expressed by a *de re* belief attribution can be determined as a function of the individual the belief is about. For these reason I will use the abstraction operator used to build up complex predicates to account for scope differences as used by Thomason & Stalnaker (1968) and formalised by Stalnaker (1977). I will assume that what property is expressed by $\hat{x}\Diamond Px$ is context dependent. To make the logic as classical as possible, I will use Van Fraassen's method of supervaluation: a sentence is true (false) in a world if it is true (false) under all admissible ways of picking out counterparts. If no admissible counterpart relation is such that there is a world in which the counterpart of the actual Caesar in that world is not dying-related to the counterpart of the death of Caesar in that world, the three cases that are problematic for Lewis's theory disappear. In particular, $\forall x,y[x = y \rightarrow \Box x = y]$ will be valid, and $\exists x,y[x= y \wedge \Diamond x \neq y]$ will be its contradictory.

What will result is a theory that is inspired most by the double-indexing account of Kaplan (1989) and Stalnaker (1978) and the counterpart theory of Stalnaker (1987b).

Syntax

The lexicon of **L** has the following ingredients:

basic symbols: $\sim$, $\wedge$, $\forall$,), (, $\hat{}$, $\Box$, ¡, †, ι, =;

---

[57] For the latter observation, see also Kripke (1972, n. 13).

a denumerable set of individual variables: $VAR_L = \{x, y, ...\}$;
individual constants: $CONST_L = \{a, b,...\}$;
the set of demonstratives: $DEM_L = \{I, you, here, now,...\}$;
for every $n \geq 0$, a denumerable set of primitive n-place predicates.

The language **L** is defined by the following definition of the terms, and complex expressions of **L**:

The set of terms of **L**, $TERM_L$, is equal to $VAR_L \cup CONST_L \cup DEM_L \cup CSTERMS_L$, where $CSTERMS_L$ is the set of complex singular terms of L to be defined later.

The set of *complex expressions* of **L** is defined as follows:

(a) Sentences

(i) If $t_1...t_n$ are terms and P an n-place predicate, then $Pt_1...t_n$ is a sentence.
(ii) If t_1 and t_2 are terms, then $t_1 = t_2$ is a sentence.
(iii) If A is a sentence, then ~A is a sentence.
(iv) If A, B are sentences, then $A \wedge B$ is a sentence.
(v) If P is a one-place predicate, then $\forall P$ is a sentence.
(vi) If A is a sentence, and t is a term, then $\Box A$ and Bel(t, A) are sentences.
(vii) If A is a sentence, then †A and ¡A are sentences.

(b) Complex predicates:

If A is a sentence and x a variable, then $\hat{x}A$ is a one-place predicate.[58]
If P is an n-place predicate, then †P and ¡P are also n-place predicates.

(c) Complex singular terms:

If P is a one-place predicate, then ιP is a complex singular term.
If t is a term, then †t and ¡t are complex singular terms.

There are no other complex expressions.

The formulae '$\exists P$' and '$\Diamond A$' will be the abbreviations of '$\sim\forall\sim P$' and '$\sim\Box\sim A$', respectively.

Semantics

Pointed models are nine-tuples $\langle W, w_0, D, *, A, R, K, C, I\rangle$, where W is a non-empty set of *worlds*, w_0 a designated element of W, representing the actual world, D is a function from W to a non-empty set such that for any two different worlds w and w', $D(w) \cap D(w') = \varnothing$. I will also denote $\cup\{D(w) | w \in W\}$ by D. * is a special object that is no element of D, A a set of *agents*, a subset of D, R a binary relation on W, K a function in $[(A \times W) \rightarrow \wp(W)]$, C a set of total functions in $[((D \cup \{*\}) \times W) \rightarrow D \cup \{*\}]$, the *counterpart functions*, and I the interpretation function. The interpretation function I meets the following conditions:

- for each individual constant c: $I(c) \in W \rightarrow D \cup \{*\}$

---

[58] If we slightly extend the language and change the interpretation rules, it would, of course, also be possible to form n-ary complex predicates, as the reader can verify.

- for any primitive n-ary predicate symbol P and any $w,w' \in W$: $I_{w,w'}(P) \subseteq (D(w'))^n$

Counterpart functions obey the following constraints:

- $\forall w \in W, c \in C, d \in D(w)$: $c_w(d) = d$
- $\forall w \in W, c \in C, d \in D$: $c_w(d) \in D(w) \cup \{*\}$
- $\forall w \in W, c \in C$: $c_w(*) = *$

If P is a primitive predicate, then $I^{\pm}_{w,w',c,g}(P) = I^{\pm}_{w,w',g}(P) = I^{\pm}_{w,w'}(P)$.

Individual terms will be interpreted in terms of a counterpart function and the object denoted by $[t]^{w,w',g}$:

$[[t]]^{w,w',c,g} = c_{w'}([t]^{w,w',g})$

The object denoted by $[t]^{w,w',g}$ is determined as follows:

$[t]^{w,w',g}$ = the utterer of t in w, if t is a *token* of *I*;[59]
= I(t)(w), if t is a constant symbol;[60]
= $[t']^{w,w',g}$, if $t = \dagger t'$ for some $t' \in TERM_L$;
= $[t']^{w,w,g}$, if $t = {}_i t'$ for some $t' \in TERM_L$;[61]
= g(t), if t is a variable;
= d, if $t = \iota P$ and $I^{+}_{w,w',g}(P) = \{d\}$;
= * otherwise.[62]

For the formal theory I assume that sentences can be true only if they are true with respect to all counterpart functions. But it would be good to distinguish two families of counterpart

---

[59] And so on for all the other demonstratives. I am cheating here a bit, because I give a token analysis only for demonstratives and not for the other expressions. I should give a token analysis for other expressions, too, that would make the interpretation rules more complicated. Instead, I just hope that you will cheat with me. If you don't, then follow Cresswell (1973, ch. 8). (In a talk given in the Sinn und Bedeutung conference in Tübingen, 1996, Manfred Kupffer took the trouble to give a token analysis semantics for a first-order language enriched with demonstratives)

[60] Note that in this way $\Delta(Pa)$ would express the same thing as $\hat{x}\Delta(Px)(a)$, if *a* is an individual constant. This doesn't mean that scope is redundant; $\Diamond(\hat{x}\Delta(Px)(\dagger a))$ is equivalent neither to $\hat{x}\Diamond\Delta(Px)(\dagger a)$, nor to $\Diamond\Delta(\hat{x}(Px)(\dagger a))$. Still, we can make scope redundant in these examples by using indexed actuality operators (see Forbes, 1985). What was expressed before by $\Diamond(\hat{x}\Delta(Px)(\dagger a))$ can also be expressed by $\Diamond_1\Delta P({}_1 a)$. If we use indexed actuality operators we have to keep track of the worlds introduced. This can be done by introducing a new sort of variable, VAR_W; assume that assignment functions are functions in $[VAR \rightarrow D] \cup [VAR_W \rightarrow W]$, say that $g[^i/_w](i) = w$, for any w, and add the following clauses:
$w,w',c,g \models \Diamond_i A$ iff $\exists w'' \in R(w')$: $w,w'',c,g[^i/_{w''}] \models A$
$w,w',c,g \models \Delta_i A$ iff $\forall w'' \in R(w')$: $w,w'',c,g[^i/_{w''}] \models A$
$w,w',c,g \models {}_i A$ iff $g(i),g(i),c,g \models A$, and
$[t]^{w,w',g} = [t']^{g(i),g(i),g}$, if $t = {}_i t'$ for some $t' \in TERM_L$.
Note that in this way we can account for the *three* readings of *Watson believes that Holmes believes that Smith's murderer is insane*, without making use of scope.

[61] Thus, '$_i$t' is the same as 'dthat(t)' in Kaplan (1989).

[62] We will see later that in this way, neither $\hat{x}\sim BALD(x)(\iota y KF(y))$ nor $\sim\hat{x}BALD(x)(\iota y KF(y))$ will have a classical truth-value if the predicate 'KF' does not denote uniquely. It is easy, however, to change the interpretation rules such that the former comes out false and the latter true. Treating definite descriptions as singular terms does not mean to adopt the Frege-Strawson theory of descriptions.

functions, one of acquaintance for the analysis of belief attributions, C_{acq}, and another for the analysis of metaphysical necessity and possibility, C_M. Both sets are subsets of C. Now we say that counterpart functions by acquaintance will become relevant only once attitude predicates are mentioned. But, of course, for the analysis of attitude attributions not all counterpart functions count. For each agent the only counterpart functions that count are the ones that represent the fact that the agent is acquainted with certain individuals. This means that for the analysis of belief attributions, the relevant counterpart functions will be world- and agent-dependent. Thus, a formula like Bel(j, A) will be satisfied in w, w', c and g, $w,w',c,g \models$ Bel(j, A), if and only if $\exists c' \in C_{acq}([j]]^{w,w',c,g,w'})$: $\forall w'' \in K([[j]]^{w,w',c,g,w'})$: $w,w'',c',g \models A$.

Proponents of situation semantics have argued that in possible worlds semantics we cannot account for the fact that in some contexts we might truly and appropriately say *Mary believes that John walks*, although this is not the case for *Mary believes that John walks and Bill talks or doesn't talk*. The reason is that Mary might have no beliefs *about* Bill at all. Unfortunately, so they claim, the embedded sentences of the two belief attributions express the same proposition according to possible world semantics, so the difference between the two sentences cannot be accounted for in this framework. But of course, once the question is one of *aboutness*, we should check in our possible world semantics whether the analysis of *de re* belief attributions can account for this difference. And it can! If Mary has no belief about Bill, there is no way in which Mary is *acquainted* with Bill, and the embedded sentence of the second clause will not express a proposition with respect to any counterpart function of Mary. The proposition expressed by the embedded sentence with respect to the presupposed belief worlds cannot be determined as a function of Bill. It follows that the belief attribution cannot be counted as being true.

To account for this intuition of aboutness we have to define a satisfaction and an anti-satisfaction relation between the quadruple <w,w',c,g> and the atomic clause Pt, and assure that with respect to the quadruple the atomic clause is neither satisfied nor anti-satisfied if $[[t]]^{w,w',c,g} = *$. Because of the way we have set up our formal language, we have to give separate interpretation rules for satisfaction and anti-satisfaction for all kinds of formulae.

I assume that if *P* is a primitive n-ary predicate, then $I^-_{w,w',c,g}(P) = (D(w'))^n - I^+_{w,w',c,g}(P)$. Of course, I hereby do not go completely classical, because * is no element of D and it is possible that $[[t]]^{w,w',c,g} = *$ for any term *t*.

The satisfaction and anti-satisfaction conditions are defined in the almost standard strong Kleene way as follows (where we leave out the superscript for the model):

$w,w',c,g \models P(t_1,...,t_n)$ iff $<[[t_1]]^{w,w',c,g},..., [[t_n]]^{w,w',c,g}> \in I^+_{w,w',c,g}(P)$
$w,w',c,g \dashv P(t_1,...,t_n)$ iff $<[[t_1]]^{w,w',c,g},..., [[t_n]]^{w,w',c,g}> \in I^-_{w,w',c,g}(P)$

$w,w',c,g \models t_1 = t_2$ iff $[[t_1]]^{w,w',c,g} = [[t_2]]^{w,w',c,g}$
$w,w',c,g \dashv t_1 = t_2$ iff $[[t_1]]^{w,w',c,g} \neq [[t_2]]^{w,w',c,g}$

$w,w',c,g \models \sim A$ iff $w,w',c,g \dashv A$
$w,w',c,g \dashv \sim A$ iff $w,w',c,g \models A$

$w,w',c,g \models A \wedge B$ iff $w,w',c,g \models A$ and $w,w',c,g \models B$
$w,w',c,g \dashv A \wedge B$ iff $w,w',c,g \dashv A$ or $w,w',c,g \dashv B$

$w,w',c,g \models \forall P$ iff $I^+_{w,w',c,g}(P) = D(w')$

$w,w',c,g \dashv \forall P$ iff $I^{-}{}_{w,w',c,g}(P) \cap D(w') \neq \varnothing$

$w,w',c,g \models \Box A$ iff $\forall w'' \in R(w')$: $w,w'',c,g \models A$
$w,w',c,g \dashv \Box A$ iff $\exists w'' \in R(w')$: $w,w'',c,g \dashv A$

$w,w',c,g \models \mathrm{Bel}(t, A)$ iff $\exists c' \in C_{acq}([[t]]^{w,w',c,g}, w'): \forall w'' \in K([[t]]^{w,w',c,g}, w')$: $w,w'',c',g \models A$
$w,w',c,g \dashv \mathrm{Bel}(t, A)$ iff $\forall c' \in C_{acq}([[t]]^{w,w',c,g}, w'): \exists w'' \in K([[t]]^{w,w',c,g}, w')$: $w,w'',c',g \dashv A$

$w,w',c,g \models \dagger A$ iff $w',w',c,g \models A$
$w,w',c,g \dashv \dagger A$ iff $w',w',c,g \dashv A$

$w,w',c,g \models ¡A$ iff $w,w,c,g \models A$
$w,w',c,g \dashv ¡A$ iff $w,w,c,g \dashv A$

If P is a complex predicate of the form $\hat{x}A$ (where $A \in \mathrm{FORM_L}$),
then (1) $I^{+}{}_{w,w',c,g}(P) = \{d \in D(w')$: $w,w',c,g[^{x}/_{d}] \models A\}$;
(2) $I^{-}{}_{w,w',c,g}(P) = \{d \in D(w')$: $w,w',c,g[^{x}/_{d}] \dashv A\}$;
(3) $I^{+}{}_{w,w',g}(P) = \{d \in D(w')$: $\forall c \in C_M$: $w,w',c,g[^{x}/_{d}] \models A\}$;

If P is a complex predicate of the form ¡Q, then $I^{\pm}{}_{w,w',c,g}(P) = I^{\pm}{}_{w,w,c,g}(Q)$;
If P is a complex predicate of the form †Q, then $I^{\pm}{}_{w,w',c,g}(P) = I^{\pm}{}_{w',w',c,g}(Q)$.

For any formula A, the absolute notions of satisfaction and anti-satisfaction are defined in terms of supervaluation: A is *satisfied* with respect to w, w' and g, $w,w',g \models A$, iff for all $c \in C_M$: $w,w',c,g \models A$, and A is *anti-satisfied* with respect to w,w' and g, $w,w',g \dashv A$, iff for all $c \in C_M$: $w,w',c,g \dashv A$.

Now we can define a notion of truth with respect to a context world *w*, and an absolute notion of truth: A is *true* in w' with respect to w, $w,w' \models A$ iff for all $g \in G$: $w,w',g \models A$, and A is *absolutely true* iff A is true in w_0 with respect to w_0.

Finally, we say that A is *valid*, $\models A$, iff A is absolutely true in all models.

It is easy to see that this semantics can account for contingent identity without giving up the intuition that objects can only be identical to themselves, and to nothing else. Consider *d*, an element of D(w), that has two representatives in *w'*, *d'* and *d''*. But how can that be if in *w'*, *d'* is not identical to *d''*? How can *d'* be *d* and *d''* be *d* if in *w'* it holds that $d' \neq d''$? The reason is (i) that according to one counterpart function, *c*, *d'* is a counterpart of *d* in *w'*, and according to another counterpart function, *c'*, *d''* is a counterpart of *d* in *w'*, although according to both counterpart functions *d* is a counterpart of *d'* in *w* ,and of *d''* in *w*, and (ii) that questions about identity are always determined from within a possible world. Although in *w'* it holds that $d' \neq d''$, from the point of view of *w*, the two represent the same individual in *w*, and thus $d' = d''$. In *w* it holds that $d = d$, $d = d''$ and $d' = d''$, because for all counterpart functions *c* it holds that $c_w(d) = c_w(d')$, $c_w(d) = c_w(d'')$, and $c_w(d') = c_w(d'')$, while in *w'* none of these equalities hold, because $c'_{w'}(d) \neq c'_{w'}(d')$, $c_{w'}(d) \neq c_{w'}(d'')$, and both $c_{w'}(d') \neq c_{w'}(d)$, and $c'_{w'}(d') \neq c'_{w'}(d'')$.

It is also easy to see that the metaphysical counterpart relations don't need to be symmetric or transitive. It is well possible for a counterpartfunction *c* in C_M that if $d \in D(w)$ and

$c_{w'}(d) = d'$, that $c_w(d') \neq d$, and that if $c_{w'}(d) = d'$ and $c_{w''}(d') = d''$, it still doesn't have to hold that $c_{w''}(d) = d''$.

Note that this semantics, just like the one given earlier, satisfies the rigidity assumption for individual constants: $t = t' \rightarrow \Box(t = t')$, and $t = t' \rightarrow Bel(a, t = t')$, for any two elements t and t' of $CONST_L$, if *a* has beliefs about the denotation of t.[63] The most important formal distinctions between this double-indexing counterpart modal logic and the quantified modal logic stated in the beginning of this chapter are that according to this semantics the clause for quantification is world dependent; that (using standard notation) the formula $\forall x \Box \exists y(y = x)$ can be false;[64] that $\exists x \exists y[x \neq y \wedge \Diamond x = y]$ is satisfiable; and that the principles of existential generalisation (EG) and universal instantiation (UI) are no longer valid. The reason that (EG) and (UI) are no longer valid is that singular terms do not have to refer to an object in the domain of quantification. More interesting is that the Free Logic versions of (EG) and (UI),

(FEG) $(A(t) \wedge E(t)) \rightarrow \exists x Ax$ (E is the existence predicate)

(FUI) $\forall x A \rightarrow (E(t) \rightarrow A(t))$, for all t

are not even valid according to the above formalism. The reason is that besides individual constants whose denotations are determined solely by the context world, there are also complex singular terms whose denotations depend on the relevant index world. That is, there is a distinction between Bel(j, P(a)) and Bel(j, P(†a)), if $a \in CONST_L$, and between Bel(j, P(ι$\hat{x}$A(x))) and $\hat{x}$Bel(j, P(x))(ι$\hat{y}$A(y)). Because universal instantiation is not valid in the above semantics, we can no longer derive the principle that any two co-referential singular terms can be substituted for each other without change in truth value, although the substitution principle of identicals (SI) is valid. In this sense, our semantics for modal logic closely resembles the proposal made in Thomason & Stalnaker (1968). The most important differences are that (i) my semantics makes use of double indexing while theirs does not, and (ii) if we reformulated their theory in terms of a counterpart theory, they would assume that the counterpart relation is an equivalence relation, while I don't. This last difference does have a formal consequence. According to the above semantics, the formula $\exists x \exists y[x \neq y \wedge \Diamond x = y]$ will be satisfiable, whereas it will not be in the semantics of Thomason & Stalnaker (1968).

Note that in our framework we can account for the intuition that the sentence *Ralph believes that Ortcutt is a spy* is true in *w* iff Ralph stands in *w* in the belief relation to the semantic value of *That Ortcutt is a spy*. The reason is that this embedded clause denotes the following semantic value: $\{<w,w',c,g> | [[Ortcutt]]^{w,w',c,g} \in I_{w,w',c,g}(Spy)\}$. With respect to world *w* and assignment *g* this is then $\{<w',c> | [[Ortcutt]]^{w,w',c,g} \in I_{w,w',c,g}(Spy)\}$. Ralph stands in *w* in the belief-relation to this semantic value iff $\exists c \in C_{acq}(Ralph,w)$: $\forall w' \in K(Ralph,w)$: $<w',c> \in \{<w',c> | [[Ortcutt]]^{w,w',c,g} \in I_{w,w',c,g}(Spy)\}$. Also, there is a semantic value that corresponds with the embedded clause of *Everybody believes that Ortcutt is a spy*, without implying that everybody must think of Ortcutt in the same way.[65] Obviously, the semantic values of *John walks* and *John*

---

[63] But it doesn't satisfy this principle for all singular terms.

[64] And thus that the Barcan formula $\forall x \Delta A \leftrightarrow \Delta \forall x A$ is no longer valid.

[65] To account for this intuition was the original motivation to represent a belief state by a pair like $<C_{acq},K>$, and to make counterpartfunctions part of indices, and was proposed in the workshop on reference and anaphora in Konstanz (1996). The structured proposition account of Cresswell & Von Stechow (1982) can account for this too, but their analysis faces a serious foundational problem: they can only account for

walks and Bill talks or doesn't talk are not the same, and so, we can account for the intuition proponents of situation semantics point to.[66]

Although Hesperus and Phosphorus, a fortnight and a period of fourteen days, and woodchucks and groundhogs actually have the same denotation and the same essential properties, we can make sense of the intuition that belief attributions like *John does not believe that Hesperus is Phosphorus, John does not believe that a fortnight is a period of fourteen days* and *John believes that no woodchuck is a groundhog* still might be true. The latter sentence, for instance, can be true if formalised, say, by Bel(j, $\sim\exists\hat{x}(Wx \wedge \dagger Gx)$). In the same way, although the H_2O-molecule that I am familiar with under the name *a* is necessary water, this is a contingent fact about our world. That is, if Putnam's twin-earth is a metaphysically accessible world, the formula $\Box$Water(a) $\wedge \sim\Box(\dagger(\text{Water})(a))$ will be true.

Of course, this second conjunct won't be true if for necessary statements we don't consider worlds where 'water' doesn't denote a stuff that doesn't have the chemical structure of H_2O. In fact, as I argued earlier, this seems the natural thing to do. In that way we can assume that the †-operator is not used inside the scope of the connectives $\Box$ and $\Diamond$. But we can systematise our account even more if we say that we do not only *never* use the †-operator inside adverbs of alethic modality, but also *always* in case of belief attributions. In fact, this generalisation of Stalnaker's approach was proposed by Haas Spohn (1994) and taken over by Zeevat (1996). We can follow them in our framework by re-defining our clause determining truth-conditions for belief attributions as follows:

$w,w',c,g \models \text{Bel}(t, A)$ iff $\exists c' \in C_{acq}([[t]]^{w,w',c,g}, w')$: $\forall w'' \in K([[t]]^{w,w',c,g}, w')$: $w'',w'',c',g \models A$

Of course, when we change our definition in this way, the distinction between the *de re* and *de dicto* reading of a belief attribution like *John believes that Hesperus is a planet* can no longer be explained in terms of whether or not the term *Hesperus* is fronted by the dagger. We are forced to explain the difference in terms of the scope of the term, and we can do so because even in belief attributions the variable, and only the variable, still is a rigid designator.[67,68]

Although in the above *semantics* the proposition expressed by a belief attribution does not depend on any particular representation of the object the speaker refers to, we might say that *pragmatically* the proposition expressed by a sentence depends on the particular

---

belief attributions like *John believes that Mary believes that the earth is flat* at the cost of assuming a whole hierarchy of different belief-predicates. My account faces no such a foundational problem.

[66] Although supervaluation theory was originally developed by Van Fraassen to account for the intuition that sentences like *Bill talks or Bill doesn't talk* are always true, this is not what we use supervaluation for.

[67] Note that if we take this line, we are forced to make belief attributions dependent on what is presupposed by the speaker who is making the attribution to account for the intuition that in normal cases the belief attribution *John believes that there is water in the bathtub* is really about what *we* call water, viz. H_2O, just like I argued for in section 1.11.

[68] It's time now to address a problem that, no doubt, worried you all along by our use of world dependent counterpartfunctions, and by the assumption I made that the domains of different worlds are mutually disjoint. These two assumptions together, gives rise to the problem how to interpret doubly embedded *de re* attributions like $\exists x|\text{Bel}(a, \text{Bel}(b, Px))|$. Fortunately, this problem can be easily solved. Let us say that if A is a formula,, f (A) gives us the sequence of variables occurring free in A. if we assume that f (A) = $\langle x_1,\ldots,x_n\rangle$, we can define f (A)(<w,w',c,g>) to be $g[^{x_1}/_{[[x_1]]^{w,w',c',g}},\ldots,^{x_n}/_{[[x_n]]^{w,w',c',g}}]$. Now we can redefine the interpretation rule for formulae of the form 'Bel(t, A)' as follows:

$w,w',c,g \models \text{Bel}(t, A)$ iff $\exists c' \in C_{acq}([[t]]^{w,w',c,g}, w'):\forall w'' \in K([[t]]^{w,w',c,g}, w'):w'',w'',c',f(A)(\langle w'',w'',c',g\rangle) \models A$

It follows now that if $g(x) = d$, and $c_{w'}(g(x)) = d'$, that $g[^{x}/_{[[x]]^{w',w',c',g}}](x) = d'$, and that if $c'_{w''}(d') = d''$, and $g[^{x}/_{[[x]]^{w',w',c',g}}] = g'$, that $g'[^{x}/_{[[x]]^{w'',w'',c'',g'}}](x) = d''$.

counterpart function that is chosen. Thus, pragmatically the proposition expressed by the embedded clause of *Ralph believes that Ortcutt is a spy* in w_0 is dependent on the relevant counterpart function: $\lambda c.\{w \in W | \forall g \in G: w_0,w,c,g \models spy(o)\}$. With respect to a relevantly different counterpart function, a different proposition would be expressed. Moreover, one belief attribution can be true and the other false in the actual world. In Kripke's example, one object of the actual world, London, will have two representatives in the worlds characterising the belief state of Pierre. Two different counterpart functions will throw out different individuals when they take the object London and a world in the belief state of Pierre as arguments. But when Pierre is expressing his beliefs by either *Londres est jolie* or *London is ugly*, he will talk about the actual London. This is because it is the actual London that is causally 'responsible' for his two representations, and what is expressed by a referential expression depends on the intentions and beliefs of the speaker. When we represent a belief state from an *external* point of view by a pair like <C, K>, where C is a set of counterpart functions, and K a set of possible worlds, we might say that the belief state from an *internal* point of view might be looked at as a pair <IC, K>, consisting of a set of possible worlds, K, and a set of individual concepts, IC, where the domain of each element of IC is K. When <C, K> is the belief state of agent *a* in *w*, we can determine IC as follows:

$$IC := \{f \in [K \rightarrow D] | \exists c \in C: \exists d \in D(w): f = \lambda w'.\ c_{w'}(d) \downharpoonleft K\}$$

If someone whose belief state can be represented from an external point of view by <C, K> expresses his beliefs and uses a word like *London*, we assume that he associates with the word a conception, modelled by an individual concept, and that the word refers in the actual world to the origin of this concept.[69] Thus, if *f* is the individual concept the speaker associates with the word *London*, his use of the world will refer in the actual world w_0 to *d* iff $\exists c \in C: f = \lambda w.\ c_w(d) \downharpoonleft K$. In the case of Pierre, there are two concepts that have the same origin. Finally, in the case of Donnellan's example of J.L. Aston-Martin, two different individuals in the actual world have only a single representative in the worlds characterising the belief state of the student. This can be accounted for by saying that two different counterpart functions, *c* and *c'*, will give the same result if applied to the two different individuals in the actual world and an arbitrary belief world of the student. We can account for the fact that in some conversational contexts the student will talk about a single person if he expresses his beliefs by his use of the name *J.L. Aston Martin* if in different conversational contexts there might be a different specific relevant counterpart function.

Note that if we represent (a) *Ralph believes that Ortcutt is a spy*, (b) *Ralph doesn't believe that Ortcutt is a spy*, and (c) *Ralph believes that Ortcutt is not a spy*, by (a') 'Bel(r, Spy(o))', (b') '~Bel(r, Spy(o))', and (c') 'Bel(r, ~Spy(o))' respectively, we would predict that (a) and (c) are true, but (b) is false in the situation sketched by Quine. The inference from (a) and (c) to (d) *Ralph believes that Ortcutt is a spy and that Ortcutt is not a spy*, however, will not be allowed.

Note that according to our semantics, we can derive (e) *Ralph believes that there is an x such that x is a spy and x is not a spy* from (d). This last inference is exactly the inference Soames (1987) wants to block by means of structured propositions. However, we will see that we don't need structured propositions to do that.

---

[69] Note that although it might be the case that the agent associates a general description with the concept, if the speaker expresses his beliefs it won't be the case that the individual he talks about will be necessarily the object in the actual world that best fits the description associated with the concept. On the other hand, if in every world *w* of the belief state of *a* the following *agent-dependent* description is associated with the concept: *The individual that I saw on Monday morning at the beach*, the agent will typically refer to the individual he saw on Monday morning at the beach if he expresses his beliefs and uses this concept.

Until now we assumed that the elements of C_{acq} are counterpartfunctions from individuals and worlds to individuals. Alternatively we might replace these sets of functions by a single function from individuals and worlds to *sets* of individuals. In that case we could say that a belief state is represented by a pair, <r,K>, consisting of such a function *r* of the latter kind and a set of worlds *K*,[70] and interpret atomic formulae as follows:

$w,w',r,g \models P(t_1,...,t_n)$ iff there is a sequence $<d_1,...,d_n>$: $\forall d_i[1 \leq i \leq n \rightarrow d_i \in r_{w'}([t_i]^{w,w',g})]$ and $<d_1,...,d_n> \in I^{+}_{w,w',r,g}(P)$.

$w,w',r,g \dashv P(t_1,...,t_n)$ iff there is a sequence $<d_1,...,d_n>$: $\forall d_i[1 \leq i \leq n \rightarrow d_i \in r_{w'}([t_i]^{w,w',g})]$ and $<d_1,...,d_n> \in I^{-}_{w,w',r,g}(P)$.[71]

In this way, different representatives of different occurrences of co-referential terms might be used to determine the truth value of a formula in a world compatible with what the agent believes, if this world is not the actual one. It is easy to see that in this way the inference from (d) to (e) will be blocked, and we can account for Soames' intuitions without making use of structured propositions. Note also that in the new semantics we can derive (d) from (a) and (b), but that the inference from (d) to (f) will be blocked, where (f) is *Ralph believes that Ortcutt is a spy and not a spy.*

It is important to see that we can define <r,K> in terms of <C,K>: Let *w* be the actual world, then r is the total function *f*: $K \times (D(w) \cup *) \rightarrow (\wp(D) \cup \{*\})$ such that :

$f(<w',d>) = \{d' \in D(w') | \exists c \in C: c_{w'}(d) = d' \ \& \ d' \neq *\}$, if $\exists c \in C: c_{w'}(d) \neq *$,
$= \{*\}$ otherwise

for any $w' \in K$ and $d \in D(w) \cup *$.

So although in the analysis of belief attributions we can do as if a belief state is represented by a pair like <r,K>, we can still in fact represent a belief state by a pair like <C,K>. Note that only from the latter kind of representation we can define the set IC of *belief objects*.

Salmon (1987) pointed to another problem for the kind of approach we implemented earlier. In our earlier formulation of counterpart theory we predict that (a) *John believes that Hesperus outweighs Hesperus*, and (b) *John believes that Hesperus is selfoutweighing* have the same meaning. Salmon (1986), however, argues that the two should be able to have a different truth value. This is something that we can account for now. If we represent the first as Bel(j, Outweigh(h, h)) and the second as Bel(j, $\hat{x}$Outweigh(x,x)(h))), we also predict a possible difference in truth value.[72] More in general, $\exists x[Bel(a, \hat{y}P(y, y)(x))]$ cannot be inferred from $\exists x,y[x = y \wedge Bel(a, P(x, y))]$. Consider for instance the case where P is $\hat{x}\hat{y}(Spy(x) \wedge \sim Spy(y))$, or even $\hat{x}\hat{y}(x \neq y)$. Also, we cannot infer Bel(a,

---

70

71 It might be suggested that if we want to analyse the sentence *Pierre believes that Londres is beautiful, and that London is ugly* as a *de re* belief attribution, it does not only count what the names refer to in the actual world, but also in what way this actual referent is named (cf. Richard (1983), and Kamp (1988)). The way the object is referred to does then also help to pick out the relevant representative of the actual London in the worlds that characterise Pierre's belief state. I would not like to propose that we should account for this intuition in a semantic way, but it is easy to see how we could. Just change the truth definition of an atomic sentence like P(t, t') as follows, if *t* and *t'* are individual constants:
$w,w',r,g \models P(t, t')$ iff there is a $d \in r_{w'}([t]^{w,w',g})$ such that $d = I(t)(w')$, and a $d' \in r_{w'}([t']^{w,w',g})$ such that $d' = I(t')(w')$, and $<d, d'> \in I^{+}_{w,w',r,g}(P)$.

72 See Heim (1993) for some discussion how this could be accounted for in a syntactically defensible way.

$\exists x[\hat{y}P(y,y)(x)])$ from $\exists x[Bel(a, P(x,x))]$. However, $\forall x,y[x = y \rightarrow \Box x = y]$ is still predicted to be valid and $\exists x,y[x = y \wedge \Diamond x \neq y]$ still its contradictory, and the death of Caesar is still predicted to be essential of Caesar, if the elements of C_M are functions from individuals and worlds to singleton sets.

Salmon (1987) argued that reflexive pronouns should be treated as predicate abstractors, and not as singular terms. Soames (1990) argued that not only reflexive pronouns should be treated as predicate abstractors, but that all kinds of anaphoric pronouns with c-commanding singular terms should be treated this way.[73] This is the best way, according to Soames, to account for the intuition that the truth of a belief attribution like *Mary believes that John loves his mother* requires Mary to believe John to be one who loves his own mother.[74] Thus, he argues that we should represent the sentence not as 'Bel(m, love(j, $\iota\hat{y}$(Mother-of(y, j)))', but as 'Bel(m, $\hat{x}$(love(x, $\iota\hat{y}$(Mother-of(y, x)))(j))' instead.[75]

Finally, Church's (1982) substitution puzzle with free variables does not arise. Church showed that under the standard assumption that variables are rigid designators, that agents stands in the belief relation to propositions modelled by sets of possible worlds, and that we don't use a counterpart theory, we can derive the counterintuitive result that $\forall x,y[Bel(a, x \neq y) \rightarrow x \neq y]$ is valid. But it is easy to see that it's apparent negation, $\exists x,y[x = y \wedge Bel(a, x \neq y)]$, is satisfiable according to our semantics without making use of structured propositions.

But isn't this a bit unfair? If we represent the information a sentence expresses with respect to a context world *w* and assignment *g* as a set of world-counterpartfunctions pairs, don't we also smuggle in some extra structure? Yes we do, but (i) we limit the extra structure needed to a very minimum,[76] and (ii) we are not committed to representationalism, the view that a belief attribution does not only express what proposition the agent believes, but also the way he believes it. That is, from the agents own point of view we can and do represent a belief state still by a set of possible worlds.

But how then can we account for self-locating beliefs? In section 1.10 we discussed the proposal of Stalnaker (1981) according to which (i) two possible worlds might be distinct, although qualitative identical, and (ii) agents can have, and only have, singular beliefs about themselves. The latter assumption seems unnatural, but I don't think that Stalnaker is committed to it. To account for self-locating beliefs we can assume that among the set of counterpart functions C of the representation of a belief state <C, K> of any agent *a*, there exists a distinguished counterpart function, c_{Self}, such that for all *w'* in K it holds that $c_{Self,w'}(a)$ is the representative of *a* in *w'*. I will assume that this counterpartfunction will obey the following constraint: $c_{Self,w'}(a)$ is an individual in *w'* that is the bearer of the same thought tokens in *w'* as *a* is in world *w*, where it is assumed that counterpart by

---

[73] Soames (1990) assumes the following definition of c-command: "A constituent *A* of a sentence *S* c-commands a non-overlapping constituent *B* of *S* iff the first branching node that dominates *A* in a constituent structure representation of *S* also dominates *B*".

[74] See also Richard (1993) who pointed to the difference between (a) *Cyril believes that John is John's father*, and (b) *Cyril believes that John is his own father*. As a matter of fact, already Lakoff (1972) pointed out that (a) *I dreamt that I was Brigitte Bardot and I kissed me* doesn't have to mean the same thing as (b) *I dreamt that I was Brigitte Bardot and I kissed myself*.

[75] Soames wonders whether the predicate-abstraction analysis should be extended to cases like (a) *Ralph believes that every man who dates Susan like her*, where the singular term antecedent of the anaphoric pronoun does not c-command it. Semantically it seems that we should, but it is in conflict with a widely assumed syntactic constraint known as *weak crossover*. As noted by Soames, if we would assume this extension, we have to explain why (a) is okay, but its quantificational counterpart (b) *Ralph believes that several men who date every woman like her* is not. See Heim (1993) for more discussion.

[76] and our treatment does not face foundational problems that all accounts using structured propositions do.

thought token goes by identity. Note that this function gives a real object as a value in every world compatible with what the agent believes, because it can be assumed that the agent himself is sure that he is thinking his thoughts. In this way we are able to determine for each agent a distinguished individual concept, which represents the way in which the agent thinks of himself. Being able to define this distinguished individual concept in terms of counterpartfunctions doesn't mean that we should analyse self-locating beliefs by means of a *de re* analysis of *I*. As explained in section 1.10, this is one way to go, but the other way of describing such situations, using diagonalisation to get rid of the 'rigidity' of the token of *I*, is still possible. By means of the second way of describing the matter, the referent of *I* in every possible world *w'* compatible with what Lingens believes will be $c_{Self,w}$(Lingens). But now we face a problem. Remember Kaplan looking in a mirror and seeing a man whose pants are on fire. At one point he wonders whether this man is he himself. How can we account for this in terms of our counterparttheory?

Until now I haven't said a lot about how we should understand the elements of $C_{acq}(a,w)$ in terms of which the counterpart relation *r* is defined should be understood. But intuitively it is reasonably clear how we should do so: the counterpartfunctions in $C_{acq}(a,w)$ mirror acquaintance, and acquaintance should be explained in causal terms. But, as stressed by Lewis (1983a), if for the analysis of *de re* belief attributions what counts are counterpartrelations by acquaintance explained in terms of causal relations, then we can only determine the counterpart by acquaintance of the individual the belief is about in a world characterising the agent's belief state when we know what the counterpart of the agent himself is in this world. This suggests that our counterpartfunctions should not be functions from individuals and worlds to individuals, but from *sequences* of individuals and worlds to sequences of individuals. The first element of this sequence must then always be the agent himself. For all counterpartfunctions *c* of the belief state <C, K> of *a* in *w* it should then hold that for all w' $\in$ K and for all $<d_1,...,d_n> \in$ dom(c(w')): $f_1(c_{w'}(<d_1,...,d_n>)) = c_{Self,w'}(a)$, when f_i is the i'th projectionfunction, and $d_1 = a$. In this way we can characterise Kaplan's beliefs when he is looking at the mirror as follows: There is a counterpartfunction *c* in Kaplan's belief state <C, K> such that *c* represents the relation x sees y as the man in the mirror. Thus, for all worlds *w'* in K it will be the case that $f_2(c_{w'}(<$Kaplan, Kaplan$>)) =$ the man $c_{Self,w'}$(Kaplan) sees in the mirror in w'. In some worlds of K, $f_2(c_{w'}(<$Kaplan, Kaplan$>))$ will be $c_{Self,w'}$(Kaplan), in others he won't.

Now that we have changed our counterpartfunctions in C, we also should change the counterpartfunctions defined in terms of it. Let *w* be the actual world, then r is the total function *f*: $K \times (D(w) \cup *) \rightarrow (\wp(D) \cup \{*\})$ such that:

$$
\begin{aligned}
f(<w',d>) \quad = \quad & \{d' \in D(w') \mid \exists c \in C: \exists <d_1,...,d_{i-1},d,d_{i+1},...,d_n> \in dom(c(w')) \ \& \\
& \qquad f_i(c_{w'}(<d_1,...,d_{i-1},d,d_{i+1},...,d_n>)) = d' \ \& \ d' \neq *\}, \\
& \text{if } \exists c \in C: \exists <d_1,...,d_{i-1},d,d_{i+1},...,d_n> \in dom(c(w')) \\
& \ \& \ f_i(c_{w'}(<d_1,...,d_{i-1},d,d_{i+1},...,d_n>)) \neq *, \\
= \quad & \{*\} \text{ otherwise}
\end{aligned}
$$

Before we conclude this chapter, we have to change our semantics in one additional way. Recall that I have argued before that the accessibility relation for determining metaphysical necessity depends on the terms used in the embedded sentence of a modal statement like $\Box A$. Until now the interpretation rules have not reflected the context-dependence of the accessibility relation. For the satisfaction case, for instance, it was stated that $w,w',c,g \models \Box A$ iff for all worlds w" metaphysically accessible from w': $w,w'',c,g \models A$. To make this metaphysical accessibility relation dependent on the terms used in the embedded sentence, let us assume that there exist a function, *f*, that for any formula *A* gives us the set of constants, demonstratives, and free variables used in *A*. With the help of this function we

can determine when w" is metaphysically accessible from w' with respect to A, w, c, and g, $w'' \in R_{w,c,g,A}(w')$, iff $w'' \in R(w')$ and $\forall t \in f(A)$: $[[t]]^{w,w'',c,g} \neq *$. Now we can restate the interpretation rules for metaphysical necessity as follows:

$w,w',c,g \models \Box A$ iff $\forall w'' \in R_{w,c,g,A}(w')$: $w,w'',c,g \models A$[77]

$w,w',c,g \not\models \Box A$ iff $\exists w'' \in R_{w,c,g,A}(w')$: $w,w'',c,g \not\models A$

We have now made the accessibility relation relevant to the interpretation of $\Box A$ dependent on A itself. This might look somewhat strange, but it is not really. According to the Lewis/Stalnaker analysis of counterfactuals, we would have to do something similar anyhow.

1.15 Conclusion

Traditionally, content was characterised by descriptive means; possible worlds and relations between possible worlds were characterised by descriptions, and what an expression refers to was assumed to be determined by the definite description the speaker, or members of his linguistic community, associated with the expression. Kripke and others have rightly criticised this picture. Possible worlds (or the relevant possibilities) and/or relations between possible worlds cannot be characterised in terms of purely general concepts, and the content of an expression should be at least partially explained in causal externalistic terms. However, we have seen that the content not only of expressions, but also of mental representations, should be explained in these causal externalistic terms. From an external point of view, belief states should be modelled by a set of world-acquaintance function pairs, and even the world information modelled by a belief state depends on the way the speaker interacts with his environment. But if this is so, we can again account for the intuition that referring is something done by speakers with terms, and not by terms themselves. What someone refers to by his use of a term depends on his intentions, which in turn depend on the content of his mental representations. But according to the causal account of content, the content of the latter representations depends on the way the speaker causally interacts with his environment. Thus, indirectly, the referents of certain terms used by the speaker are determined by the causal relations the speaker bears to the world. This, as I will argue in the next chapter, is the case not only for proper names, but also for most (other) uses of anaphoric expressions. And, just as agents might be unclear about what the referent of a proper name is because they are unclear about the origin of the relevant referential chain, agents might be unclear about what the referent of a pronoun is because they are unclear about the causal origin of the relevant anaphoric chain. In this chapter, I have argued that the first kind of unclearness for referential chains should be modelled by diagonalisation. In the next chapter I will argue that things are the same for anaphoric chains.

In § 1.14, we have come to the conclusion that to account for *de re* belief attributions, we should model a belief state of an agent in world *w* from an external point of view by a pair <C, K>, where K is a set of possible worlds consistent with what the agent believes in *w*, and C is a set of functions from individuals in *w* and worlds in K to individuals in these worlds. In this way we can represent ways in which agents have beliefs about individuals. This representation of belief states is close to the representation of anchored beliefs in Kamp (1990). Then the question arises in what ways agents can form beliefs about individuals, or in terms of Kamp (1990), which relations give rise to anchored beliefs? Kamp suggests three such relations: visual perception, memory, and the forming of a new belief in response to an utterance which contains a direct referential expression. I agree that in all those three ways agents can form beliefs about individuals, but I also think that it is much easier to form beliefs about individuals by means of communication. We don't have

---

[77] You might argue that this interpretation rule has an unfortunate consequence: $\forall x \Delta \exists y(x = y)$ will be valid again.

to accept assertions in which a, what is traditionally called, direct referential expression occurs; normal indefinites and pronouns will do. In the next chapter I will argue that presupposition states should be represented in basically the same way as belief states, representing the information that participants in a conversation have of the individuals whom the conversation is about.

Chapter 2

Referential and Descriptive pronouns

2.1 Introduction

One of the things I defended in the first chapter was that the meaning and content of linguistic expressions should be explained in terms of the intentions, beliefs and conventions of language users. This is in particular the case for pronouns, the expressions I will concentrate on in this chapter. What a pronoun refers to depends on the intention of the speaker. But intention comes in two forms: we have *specific* and *general* intention (cf. Kripke, 1977). In this chapter I want to argue that for most uses of pronouns it is the specific intention that counts, only in exceptional cases it is the general intention. Stated differently, I claim in this chapter that pronouns are normally *referentially* used, and only sometimes *descriptively*. Normally, speakers intend to refer with their use of a pronoun to the *speaker's referent* associated with the indefinite that figures as its syntactic antecedent, sometimes to the referent of the definite description recoverable from the sentence in which its syntactic antecedent occurs, if this description has a unique referent at all. In this chapter I will show that using the *two-dimensional* theory of reference and the Stalnakerian (1978) *diagonalisation* strategy, the referential analysis of pronouns can be pushed further than many have supposed. The insight of the two-dimensional theory of reference was that it is useful to separate facts about the conversation's subject matter from facts about the conversational situation itself. By diagonalisation it is possible to model what is going on if an expression is referential used, although it is unclear what the speaker actually referred to with the expression. The resulting theory will closely resemble popular modern theories of anaphora like Discourse Representation Theory (Kamp, 1981), File Change Semantics (Heim, 1982), and their more recent dynamic counterparts due to Groenendijk & Stokhof (1991) and Dekker (1993) in that indefinites introduce objects to the discourse that become available for reference by pronouns. But the approach I will argue for also differs from those popular theories on some crucial points. For one thing, in contrast to the popular theories of anaphora, the referential approach I favour allows for a natural explanation why we also need descriptive pronouns. The E-type approach of Evans (1977), Cooper (1979) and Neale (1990) is not seen as a competing approach to the referential one. I will argue that, instead, the referential and E-type approaches are natural complements of each other. Whether a *singular pronoun* is referentially or descriptively used, there is always an assumption of *uniqueness* involved. Either the pronoun is referentially used, and it refers to the unique speaker's referent in the relevant possibility, or the pronoun is descriptively used, and it refers to the unique individual of the world of the possibility that satisfies the associated description.

Additionally, I also want to argue for a *salience* based analysis of anaphoric reference. In DRT/FCS the assumption is made that there exists an absolute distinction between objects that are and objects that are not available for reference for short definite expressions, and that indefinites can transfer objects from the latter group to the former. That certainly was a useful hypothesis, but something more seems to be required to account for the fact that some of the objects available for reference can be more easily referred back to than others. For some we just need a pronoun, while for others we need a definite expression with more descriptive content. I will argue that a more sophisticated version of the theory should follow Lewis (1979b) by ordering the possible objects available for reference with respect to their salience. A definite noun phrase like *the man* will then refer in a possibility to the most salient man in this possibility relative to this conversation.

I first discuss some classical approaches towards anaphora, and formulate dynamic semantics as it is standard by now. Then I argue that most pronouns are referentially used, and account formally for referential pronouns in a dynamic framework. But sometimes a singular pronoun used in the 'main' context can be appropriately used although it does not refer to the speaker's referent of the indefinite. This will motivate postulating the existence

of descriptive pronouns. I will implement these pronouns in dynamic semantics in a systematic way. Next, I will argue for a specific way in which epistemic modals can be 'anaphorically' dependent on each other. Then I will sketch a way in which *functional pronouns*, needed to account for Karttunen's (1969) notorious paycheque examples, can be accounted for in our dynamic framework in such a way that the uniqueness constraint on singular pronouns will still be satisfied. In the following section I will prepare the grounds for my analysis of presuppositions in quantified contexts in chapter 4. First, I will argue that not only pronouns, but also quantifiers and even indefinites should be treated as anaphoric expressions in a weak sense. Second, I propose a formalisation of the intuition that an anaphoric expression picks up the most salient object that satisfies the descriptive content of the expression. For that reason I will order the possible objects available for reference with respect to their salience. In the final section it is argued that if both speaker's reference and the salience order are world-dependent, we might formulate dynamic semantics that accounts for referential pronouns in terms of possible worlds only.

2.2 Some classical approaches towards anaphora

According to the common understanding of scholastic approaches towards indefinites and anaphora, pronouns can refer back to indefinites because indefinites are referential expressions. The indefinite refers to that object that the speaker intends to refer to by the use of the indefinite. Moreover, if a speaker uses in his utterance a referential expression, the proposition expressed by this utterance is object dependent. Geach (1962) has criticised this account. If John intends to refer to *a* by his use of the indefinite *an S*, and wants to say of *a* that he is *P*, although *a* is not, John is not saying something false when he claims *An S is P*, according to Geach, if there actually is an S that is P. Not to make such a prediction, Geach proposed it is better to represent an assertion like *An S is P* semantically simply by an existential formula, $\exists x[Sx \wedge Px]$. The specific/unspecific distinction belongs to pragmatics, which should be kept separate from semantics. To handle pronouns, we should follow Quine's insight, and treat them as bound variables. A sequence of the form *Some S is P. It is Q* should according to him be translated as $\exists x[Sx \wedge Px \wedge Qx]$. But there are well known problems with this latter assumption. First, it leads to the unnatural assumption that we can only interpret a sentence with an indefinite or other anaphoric initiator at the end of the whole discourse: incrementallity is given up. Second, if we want to interpret the pronouns in a donkey sentence *If a farmer owns a donkey, he beats it* as bound variables, it seems we have to represent the indefinites in the antecedent as universal quantifiers to get the truth conditions right. But then it seems we have to give up compositionality. We cannot treat indefinites in all contexts in the same way. Finally, sometimes we cannot even get the truth conditions right by assuming that all pronouns should be treated as bound variables. This was shown by Gareth Evans (1977). Evans convincingly argued that sometimes by a use of a pronoun we denote *all* the relevant objects by which the antecedent sentence is verified. Thus, in a sequence of the form *Some S are P. They are Q*, the pronoun *they* is going proxy for the description *(all) the S such that P*.[78] Such pronouns he called *E-type* pronouns, but I will sometimes also call them *descriptive pronouns*. The existence of E-type pronouns was argued for by the following kind of example:

(1) Tom owned some sheep and Harry vaccinated them.

According to a Geachian analysis of the sentence we only learn that Harry vaccinated some sheep that Tom owned if we accept what is expressed by the sentence, but what we seem to learn in fact is that Harry vaccinated *all* sheep that Tom owned. The latter reading is predicted if the pronoun *them* is analysed as an E-type pronoun.

[78] Evans (1977) claimed that the pronoun *rigidly refers to* (all) the S such that P. See Neale (1990) for a motivation for the interpretation I have chosen. I will give some additional motivation later. Still, I agree with Evans's claim that most unbound pronouns are referring expressions, however, I will argue that these pronouns are not E-type pronouns.

I think it is undeniable that E-type pronouns do exist, but that doesn't mean that all pronouns are E-type pronouns. First, there is the obvious reason. The pronouns occurring in

(2) *Every man* loves *his* cat, and
(3) *Each woman* liked the man who gave *her* a rose.

seem to function like the bound variables of quantification theory. Indeed, since Evans (1977), proponents of the E-type approach normally make a distinction between *bound* and *unbound* pronouns, claim that such a distinction can be made on purely syntactic grounds, and propose that only unbound pronouns should be treated as E-type pronouns. A pronoun *P* is a bound pronoun, and treated as a bound variable, roughly if it is anaphoric on, and thus bound by, a quantifier Q, only if *P* is located inside the smallest clause containing Q (Neale, 1990, p. 171).[79]

However, if we use the term *unbound pronoun* in the above sense, it seems that not even all unbound pronouns go proxy for the definite or universal noun phrase recoverable from the antecedent clause, and thus should be treated as E-type pronouns. Consider the following example due to Dekker (1994):[80]

(4) Yesterday, John met some girls. They invited him to their place

In this case, we don't want to say that *they* needs to stand for all girls John met yesterday. If we want to say that the pronoun is going proxy for a description recoverable from its antecedent, the relevant description should not be definite or universal, but *indefinite*.[81] The description would be *Some girls that John met yesterday*. To treat the pronoun as an abbreviation of an indefinite description also seems to be needed to get the right reading of a sentence like

(5) Socrates owned a dog, and it bit Socrates.

It seems that (5) can be true if there was a dog that Socrates owned and bit him, although at the same time there was also another dog that he owned which did not bite him. But claiming that the pronoun is an abbreviation of an *indefinite* description would be very implausible. Pronouns are definite expressions:

> 'It' [is] a definite singular term whether its antecedent is or not. 'He', 'she', and 'it' are definite singular terms on a par with 'that lion' and 'the lion' [...] The three compound sentences 'I saw a lion and you saw that lion', 'I saw a lion and you saw the lion', and 'I saw a lion and you saw it' are interchangeable. Such use of a definite singular term dependently upon an indefinite antecedent [...] makes no distinction between a pronoun such as 'it' and a singular description such as 'the lion' (Quine, 1960, p. 113)

Should we thus treat all unbound pronouns as E-type pronouns after all? Maybe there is a way to explain away the uniqueness prediction that is in some cases apparently unwelcome if singular anaphora are treated as E-type pronouns.[82] Maybe, but prospects look dim. First, it doesn't seem to be a natural strategy to explain away 'apparent' counterexamples to the uniqueness assumption by assuming that the domain of quantification is always selected in such a way that the uniqueness effect is reached after all. Second, sometimes domain restriction even doesn't help. This is shown by donkeys in bishop clothing's:

---

[79] For a more specific syntactic characterisation, see Evans (1977) and Neale (1990). In the later discussion on epistemic *might* I will argue that there is indeed something to the distinction between bound and unbound pronouns as syntactically characterised by these authors.

[80] For similar examples, see Sommers (1982) and Kamp & Reyle (1993).

[81] See Van der Does (1994), and Meyer Viol (1995).

[82] For early discussion, see Evans (1977); see Neale (1990) and Heim (1990) for some more recent ones.

(6) If a bishop meets another man, he blesses him. (Heim, 1990)[83]

If E-type pronouns are treated as *definite* descriptions, it seems to be impossible to select the domain in the correct way. As argued above, giving up the assumption that pronouns are definite expressions doesn't seem to be natural.

But if a singular pronoun cannot be treated as the definite description that (in extensional contexts) refers to (all) *the* object(s) that verify the antecedent sentence, how then can a pronoun be treated as a definite expression?

The answer of Kamp (1981), Heim (1982), and more recent proponents of dynamic semantics like Groenendijk & Stokhof (1991) and Dekker (1993) is familiar by now: treat anaphoric pronouns simply as bound variables, interpret indefinites dynamically such that they introduce new objects available for reference, and assure that in case of negation a universal quantification over assignment functions or sequences of individuals is involved. Anaphoric pronouns can be treated as definite noun phrases, because the possibilities with respect to which the pronouns are interpreted are made more fine-grained entities then possible worlds, namely, world-assignment pairs. From now on I denote all dynamic theories simply by CCT, for *Context Change Theory*. I will state now standard CCT as it is formulated in Dekker (1993). In this formalisation I assume that all terms are variables.

2.3 Context Change Theory

Syntax

The syntax of the language **L** is the same as that of standard first-order predicate logic without individual constants.

The lexicon of **L** has the following ingredients:

basic symbols: $\sim, \wedge, \vee, \rightarrow, \exists, \forall,), (, =$;
individual variables: $VAR_L = \{x_1, x_2, ...\}$;
for every $n \geq 0$, the set of n-place predicate constants: $PRED^n_L = \{P^n_1, P^n_2, ...\}$

The language **L** is defined by the following definition of the terms and formulae of **L**:

The set of terms of **L**, $TERM_L$, is equal to VAR_L

The set of formulae of **L**, $FORM_L$, is the smallest set such that:

(i) if $t_1...t_n \in TERM_L$ and $P \in PRED^n_L$, then $Pt_1...t_n \in FORM_L$;
(ii) if $t_1, t_2 \in TERM_L$, then $t_1 = t_2 \in FORM_L$;
(iii) if $A \in FORM_L$, then $\sim A \in FORM_L$;
(iv) if $A, B \in FORM_L$, then $A \wedge B \in FORM_L$;
(v) if $A \in FORM_L$ and $x \in VAR_L$, then $\exists xA \in FORM_L$.

Semantics

[83] Attributed to Kamp and Van Eijck.

Models are triples <D, W, I>, where D is a non-empty set of objects, W a non-empty set of possible worlds, and I the intensional interpretation function that maps n-ary relations to a function from worlds to sets of n-tuples of objects.

The set G of *partial assignments* associated with D and **L** is $\cup\{D^X \mid X \subseteq VAR_L\}$

An *information state* S with domain X is a set of assignment-world pairs ($S \subseteq G \times W$) such that for all <g,w> that are elements of S it holds that X is the domain of *g*. I will say that in these cases X is the domain of S, D(S) = X. I will use the following notational conventions with assignments *g* and *h*, objects *d*, variables *x* and *y*, and worlds *w*, where $x \notin dom(g)$ and for no $<g,w> \in S$: $x \in dom(g)$:

$g[x]h$ iff $dom(h) = dom(g) \cup \{x\}$ & $\forall y \in dom(h)[y \neq x \rightarrow h(y) = g(y)]$

$S[x] := \{<h,w> \mid \exists g:<g,w> \in S \ \& \ g[x]h\}$

$S[x:= d] := \{<h,w> \mid \exists g:<g,w> \in S \ \& \ g[x]h \ \& \ h(x) = d\}$

The elements of ($G \times W$) are ordered by $\leq$: $<g,w> \leq <h,w'>$ iff $w = w'$ and $g \subseteq h$. This ordering relation carries over to information states S and S': $S \leq S'$ iff for every $\alpha \in S$: there is an $\alpha' \in S'$: $\alpha \leq \alpha'$.

For the interpretation rule of negation I introduce $\alpha \ll S$, saying that α has an *extension* in S, which is the case iff $D(\{\alpha\}) \subseteq D(S)$ & $\exists \alpha' \in S: \alpha \leq \alpha'$. *Subtracting* state S' from state S, S - S', will leave us with those elements of S that have no extension in S': $S - S' = \{\alpha \in S \mid \sim(\alpha \ll S')\}$.

The notation 'G(S)' will be used to give us the set of assignment functions in S:
$G(S) = \{g \in G \mid \exists w \in W: <g,w> \in S\}$

If <g,v> is an assignment-world pair, w(<g,v>) = v.

Now I can give a recursive definition of the context change potential [[A]] [$\subseteq \wp(G \times W) \times \wp(G \times W)$] of formulae A of **L**:

(1a) $[[Px_1...x_n]](S) = \{\alpha \in S \mid <\|x_1\|^\alpha,...,\|x_n\|^\alpha> \in I_{w(\alpha)}(P)\}$,
if $\forall x_i: 1 \leq i \leq n: \forall \alpha \in S: \|x_i\|^\alpha$ is defined, undefined otherwise

(1b) $[[x_1 = x_2]](S) = \{\alpha \in S \mid \|x_1\|^\alpha = \|x_2\|^\alpha\}$,
if $\forall x_i: 1 \leq i \leq 2: \forall \alpha \in S: \|x_i\|^\alpha$ is defined, undefined otherwise

The (static) term-evaluation used in (1a) and (1b) is defined by:

$\|x\|^{g,w} = g(x)$, if $x \in dom(g)$, undefined otherwise

Given the induction step I assume that [[A]] (S) and [[B]] (S) have already been defined (for given formulae A and B and information states S) and give the following:

(2) $[[\sim A]](S) = S - [[A]](S)$
$= \{\alpha \in S \mid \sim\exists \alpha'[\alpha \leq \alpha' \ \& \ \alpha' \in [[A]](S)]\}$

(3) $[[A \wedge B]](S) = [[B]]([[A]](S))$

(4) $[[\exists x]](S) = \{<h,w> \mid \exists g:<g,w> \in S \ \&\ g[x]h\}$ $(= S[x])$
if $\forall g \in G(S)$: $x \notin dom(g)$, undefined otherwise

Disjunction and implication can be treated syncategorematically, by having '$(A \vee B)$' and '$(A \rightarrow B)$' stand for '$\sim(\sim A \wedge \sim B)$' and '$\sim(A \wedge \sim B)$' respectively. A formula like '$\exists xA$' is analysed as the conjunction of '$\exists x$' with 'A', and '$\forall xA$' will be taken as an abbreviation for '$\sim\exists x\sim A$'.

To define the different notions of acceptability and truth, we first need to define when information state S is a substate of state S': S is a *substate* of S', $S \copyright S'$, iff for every $\alpha \in S$: there is an $\alpha' \in S'$: $\alpha \leq \alpha'$

Now I can define the most important semantic concepts. A formula A is *acceptable in S*, $S \models_d A$ iff S is a substate of $[[A]](S)$, in the sense that every $\alpha \in S$ can be extended to an $\alpha' \in [[A]](S)$ such that $\alpha \leq \alpha'$. A is *accepted in S*, $S \models_S A$ iff $S = [[A]](S)$. A is *true in* $<g,w>$, $<g,w> \models A$ iff $\{<g,w>\} \models_d A$. A is *true in w* with respect to S, $w \models^S A$, iff there is a g: $<g,w> \in S$ and $<g,w> \models A$. A *entails B*, $A \models_{d/S} B$, iff for all S: $[[A]](S) \models_{d/S} B$.

2.4 Anaphoric pronouns as referential expressions

Context Change Theory as stated above is empirically quite successful in accounting for anaphoric relationships across sentential boundaries. I think, however, that the theory faces a conceptual problem. In order to account for anaphoric dependencies in a compositional way, the elements of information states are not possible worlds, but world-assignment pairs, instead. As a result, possibilities of information states are more fine-grained entities than possible worlds, and a CCT-information state contains more than just truth-conditional content. But then the question arises what this extra content can be, what this extra fine-grainedness of the possibilities represents.

To focus the discussion, let us introduce the notion of a *subject* as defined by Dekker (1993). Consider only a model with one possible world. In that case a CCT-information state, S, might be modelled by a set of assignment functions. Normally, assignment functions are seen as functions from elements of VAR to D. But as noted by Janssen (1986) we might as well say, instead, that variables are functions from elements of S to D. So, we can define the information of S associated with variable *x* as:

$[x]_S :=$ the function $f \in [S \rightarrow D]$ such that $\forall \alpha \in S$: $f(\alpha) = \alpha(x)$

With respect to information state S no information is lost. Dekker (1993) calls such a function the *subject* of S stored under *x*.[84] Our question what the extra content is that CCT-information states contain can now be rephrased as asking what subjects represent, if they represent anything at all. In DRT-terms, the question is what *discourse referents* stand for.

This question has been discussed in two recent papers of Zimmermann (1995) and Dekker (1996).[85] Both seek to explain the status of subjects in a non-representational way; subjects of information states should represent something about the actual world, they should not

---

[84] Of course, when we consider models with more possible worlds, subjects can be defined in the same way.

[85] See also Kamp (1990) and Spohn (1997).

just be a tool for determining the truth-conditions of sentences that are interpreted with respect to this information state. Otherwise proponents of CCT would be committed to representationalism.

Zimmermann discusses and rejects the most straightforward *meta-discourse* solution to our problem. According to this solution, the presence of a discourse referent in an information state represents the information that a certain noun phrase has been used in the discourse, one that can be taken as syntactic antecedent of a pronoun. One might think that the reason for a non-representationalist to reject this proposal is that this solution assumes that a CCT-information state would contain more than just truth-conditional information about the subject matter of conversation. But as already noted by Stalnaker (1978), and stressed in Stalnaker (1996), there is nothing representational about this assumption. An assertion that is accepted changes what is presupposed in two ways: not only will the content of what is asserted become common background, but the fact that the speaker asserted something by his use of a certain sentence does as well. Thus, the context is not just incremented with information about facts relevant to the conversation's subject matter, but also with information about the conversation itself.[86] Information about the conversation is information about the world, and as long as a theory makes use of information states that contain only information about facts of the world, the theory is not committed to representationalism.

The meta-discourse solution is not problematic because it assumes that information states contain information about the discourse itself. The solution is rejected by Zimmermann because it doesn't explain by itself why pronouns normally need explicitly mentioned indefinites they can take as their syntactic antecedents. The meta-discourse solution only ascribes to discourse referents some descriptive content like *introduced by the indefinite 'an S'*, and does not affect their status as existential quantifiers. But the information states in CCT should not only contain the information that a certain noun phrase has been used, it should also contain enough information in terms of which we can explain in a natural way why a pronoun can take a certain indefinite as antecedent. Saying that a pronoun can be used because it is common background that an indefinite was used before, raises the question what it is about the use of an indefinite that makes it possible to appropriately use a pronoun later. What can this extra information be that discourse referents, or subjects, represent?

Zimmermann (1995) proposes that subjects represent *sources* of information. But, then, what is a source? Is it the token of the indefinite by which the subject is introduced, or the individual that is causally 'responsible' for the speaker's use of the indefinite by which the subject is introduced? It seems that neither will quite do. We have argued above that a subject should not just represent the information that a certain noun phrase has been used, and as argued by Zimmermann (1995), if it is a *fata morgana* that was responsible for the speaker's use of the indefinite, we don't want to claim that the subject introduced by this indefinite is a representation of this fata morgana.

I want to propose that a subject represents the presumed speaker's referent of the use of the indefinite (or other act of the speaker) by which it is formally introduced. Normally, a subject represents for a hearer the unique thing or individual the speaker has in mind about which he intends to speak. The speaker's audience will normally not be able to recognise which individual the speaker intends to talk about from the context and the sentence in which he introduced the subject. It follows that it becomes appropriate for the hearers to ask for more information about this individual (cf. Donnellan, 1978). With Dekker (1996) we can say that normally the participants in the conversation also want to gain more information about this individual.

---

[86] This, of course, is needed to determine the referents of anaphorically used descriptions as *the former*, and *the latter*.

I said that the speaker's referent is "the individual the speaker had in mind". But this latter notion should be taken with care. It need not be the case that the speaker has an impersonal definite description in mind that uniquely fits an individual. That would indeed be unexpected given the discussion in chapter 1. What is normally meant by the "individual in mind" is the object that is the dominant source of a particular body of information relevant to the speakers use of the indefinite on a particular occasion. It can be that some information the speaker has about this object is wrong, and if he is expressing his beliefs he might be saying something wrong about the individual he is talking about.

Earlier in this chapter I mentioned the approach - the E-type approach - according to which every pronoun should be treated as a descriptive pronoun, and discussed some at least superficially problematic consequences of it. It is easy to see, however, that the approach cannot account for all anaphoric dependencies across sentential boundaries. This is due to the phenomenon of *pronominal contradiction*. When John asserts *A man is walking in the park*, Mary can react by saying *It's not a man, it's a woman.* By treating the pronoun as an abbreviation of the description *the man who is walking in the park*, Mary's remark would be trivially false. In these cases the pronoun appears to be used referentially, referring to the individual John had in mind by his use of the indefinite.

Caces of pronominal contradiction suggests that at least sometimes pronouns are used referentially, referring back to the specific object the speaker had in mind by his use of the indefinite antecedent. The following phenomenon suggests that pronouns are in general used this way.[87] If *a* says *A man called me up yesterday*, it would be odd for *a* to reply to *b*'s question *Did he have a gravel voice?* by saying *That depends, if he called up in the morning he did, if he called up in the afternoon, he did not* if in fact two men called up *a* yesterday. It not easy to see how this phenomenon can be explained if it is assumed that pronouns should simply be treated as variables bounded by dynamic existential quantifiers. On the other hand, a natural explanation of it can be given if it is assumed that pronouns are in general referentially used.

Indeed, something like this has been proposed by Strawson (1952), Chastain (1975), Kripke (1977), Donnellan (1978), Lewis (1979b), Fodor & Sag (1982), and Stalnaker (1996). A singular pronoun sometimes can take an indefinite description as syntactic antecedent, because the speaker had a specific individual in mind to which he intended to refer with the indefinite. How should we account for the suggestion that anaphora are sometimes treated as referential expressions?

According to the two-dimensional theory of reference, a sentence determines a proposition, a function from worlds to truth-values, relative to a context. It is commonly assumed that to determine the referents of referential expressions, one can represent a context by an n-tuple of objects, containing a speaker, a hearer, a place, a time, and some salient objects the speaker can refer to by the use of deictic pronouns or demonstratively used descriptions.

There seems to be a good reason why a context should be represented by a single tuple of objects that are available for reference by referential expressions such that the elements of this tuple are *common ground* between speaker and hearer. The reason is a Gricean one: speakers ought to assume that hearers have enough information to determine what proposition is asserted by the speaker. If the hearer fails to recognise what object is referred to by a referentially used expression, then the hearer cannot determine what proposition is expressed by the speaker, and thus a conversational maxim is violated. It seems to follow that if some anaphorically used pronoun is treated as a referential expression, the speaker has to presuppose that the hearer can recognise to what object the speaker was intending to refer by the pronoun in order to understand what is said by the sentence in which the pronoun occurs. Unfortunately, this is generally not the case for the anaphoric pronouns treated by CCT. On the other hand, this is not even the case for the demonstratively used

[87] This example came up in a discussion with Paul Dekker and Ede Zimmermann.

pronouns, where the speaker not only has a particular object in mind, but also assumes that the hearers know which object this is. But how, then, can communication be successful if the hearer cannot tell to which object the speaker was intending to refer with the pronoun?

Although in the ideal case a referential expression is used only when it is clear to the hearer what the expression refers to, it is clear that ideal conditions do not always obtain. If the speaker says something and the hearer disagrees, there might be two reasons for this disagreement. First, the hearer has understood what the speaker has said, but he disagreed with the speaker about the facts. Second, speaker and hearer might agree about these facts, but still disagree, because the hearer has thought that the speaker has said something different from what he has actually intended to say. The latter might be the case when the speaker uses a referential expression. These two different reasons for disagreement can be accounted for in the two-dimensional theory of reference from Stalnaker (1970b) and Kaplan (1989) by means of the Stalnakerian (1978) diagonalisation strategy. The reason is that in the two-dimensional theory a distinction is made between two kinds of facts: (i) facts about the subject matter of conversation or thought, and (ii) facts about the conversational situation itself, and more general linguistic and speech conventions.[88] In a simple example of Stalnaker (1978), *a* is talking to *b* and *c*. Although *a* and *b* believe that *c* is a fool, and *a* intends to refer to *c* by the use of the demonstrative pronoun *you* in his assertion *You are a fool*, *b* might disagree with *a* because he believes that *a* was referring to him, *b*, by the use of the pronoun, and *b* does not think of himself as a fool. If *b* were a cautious hearer, he would not have come so quickly to the wrong conclusion that *a* was referring to him. He would think of himself as being unclear about what proposition *a* has intended to assert by saying *You are a fool*. In this simple situation we can think of a reference-context as a possible referent of the demonstrative pronoun *you*. Clearly, there are two possible referents, *b* and *c*. If *b* were a cautious hearer, he would represent what has been said by *a* as a function from a reference-context to the proposition expressed by *You are a fool*: in this reference-context $\{\{w \in W | d \text{ is a fool in } w\} | d = b \text{ or } d = c\}$.[89] Of course, this function from reference-contexts to propositions is formally a Kaplanian (1989) *character* or Stalnakerian (1978) *propositional concept*. In the above conversational situation, a sentence like *You are a fool* can express two kinds of propositions.[90]

If *c* is a reference-context, the *horizontal proposition* expressed by *A* with respect to *c* can be denoted by [A](c), and is determined as follows: $\{w' \in W | w' \in [A](c)\}$. Although it is normally the horizontal proposition with respect to the actual reference-context that the speaker intends to express, we have seen that a hearer sometimes doesn't know which one this is. If the hearer is cautious, the information that he can receive is different.

Until now a context has been taken to represent only the information available to interpret context-dependent utterances, but it should also contain the information that is accepted by speaker and hearer about the *subject matter* of the conversation. It is this information that speakers try to influence by making assertions. To combine these two kinds of information

---

[88] It is widely assumed that the crucial insight of dynamic semantics is that the meaning of a sentence should be equated with its context change potential, not with its truth conditions. But it is more appropriate to say, I believe, that the crucial insight of dynamic semantics is that one receives more information from an assertion than what it says about the subject matter of conversation; one receives also information about the conversational situation itself.

[89] In this chapter I ignore counterpart functions.

[90] In a sense the character or propositional concept of a sentence is just an underspecified representation of what is said by the sentence. But only up to a certain degree; it does not represent semantic or syntactic underspecification. Where underspecified representations can account for lexical ambiguities and ambiguities of scope, this is not the case for Kaplanian characters. Scope is no issue, and although *you* might *refer* to different individuals in different reference-contexts, Kaplan (1989) taught us that the reference of a demonstrative is not all there is to its *meaning*. The demonstrative *you* will denote a hearer in all contexts of interpretation, it has an unambiguous character. A word as *bank*, on the other hand, has an ambiguous character.

that a context contains, we should now represent a context C as the set of reference-context/index pairs in which everything accepted in the conversation is made true.[91] If any element of C might, as far as the hearer can tell, be the actual reference-context/index pair, he might update this information state after accepting the utterance by eliminating any reference-context/index pair <c,w> in C in which what is expressed in *c* is false in *w*. This new information state is {<c,w> ∈ C| w ∈ [A](c)} and is what Stalnaker (1978) has called *the diagonal* of *A* with respect to C.

This latter way of updating a context looks very much like what is going on in the dynamic semantics stated earlier. And indeed, I want to propose that a subject represents the information the hearer has about a certain speaker's referent; the diagonalised speaker's referent. But, then, the question arises whether we ever need more than just the diagonal, do we ever need the whole propositional concept?

Obviously we do, it seems, because the speaker normally intends to express the horizontal proposition determined by the sentence. But proponents of the standard dynamic account will not be easily impressed. Look at a case where the speaker uses a demonstrative pronoun. In an assertive utterance of *You are sick* the speaker certainly intends to express the horizontal and *object dependent* proposition expressed by the sentence, a proposition that says something *about* a particular object. Still, looking only at the diagonal does not really seem problematic. But isn't it true that the diagonal proposition expressed by the above sentence is context-, and thus object-independent? Yes, in principle this diagonal proposition is object-independent, but what is relevant is always the diagonal proposition expressed with respect to a special context. If all reference-context/index pairs were relevant, the above claim wouldn't say much more than *the person to whom the speaker is speaking is sick*. In most conversational contexts, however, it is pretty clear to the participants of the conversation who the speaker and addressee are: the same individuals in all reference-contexts of the context. It follows that by the above claim the same object-dependent proposition would be determined in all the relevant reference-contexts, and thus the diagonal of the relevant propositional concept would be object-dependent, too.

Still, I agree with Stalnaker (1970b) that we need the whole propositional concept. Reference-context/index pairs should not be merged into primitive points of reference such that propositions are considered to be functions from reference points to truth values.[92] The horizontal propositions expressed are of some independent interest, and to bring that out there has to be a functional difference between reference-context and index. One reason we need sometimes the full propositional concept, is that we want to be able to make a distinction between the *a priori* and the *necessary*. If we always look only at the diagonal, in a sense we always look whether what is said is *a priori* true. We only look at things from an epistemic point of view. But sometimes we want to know whether what is said is, for instance, *physically necessary* true or not, as in *I didn't have to be here you know* (Stalnaker, 1970b). In these cases the horizontal proposition is relevant. Similarly, a sentence of the form *It may be that A* can express that A is consistent with what is presupposed, but can also express the modal proposition that A is consistent with some suitable chosen set of (physical, logical, ethical, ...) law sentences.[93] In the latter case, it is again the horizontal proposition that is relevant.

This horizontal proposition should normally be determined with respect to the worlds compatible with what is presupposed about the subject matter of conversation. This can be shown by another reason why we need to be able to determine the whole propositional concept: *referential disambiguation*. It is obvious that when the speaker uses a

---

[91] For simplicity I will assume that indices are possible worlds.

[92] See also Lewis (1980)

[93] For this reason, Stalnaker (1970b) says that sentences of the form *It may be that A* are *pragmatically ambiguous*.

demonstrative pronoun like *you* in a sentence like *You are a fool*, he intends to say something about a specific individual, and thus intends to communicate the horizontal proposition. Only because it is sometimes unclear to the hearer what the speaker has intended to say does the diagonal become relevant. Let us assume that I uttered *I will see you at 10 o'clock tomorrow* in a conversation with Ede, Paul and Tim. Suppose that it is common knowledge among us that Paul is taking the night train from Stuttgart to Amsterdam this very evening. In that case, even if my pointing has not been very clear, it will be clear to the three hearers what I did not intend to say: namely, that I will see Paul tomorrow at 10 o'clock. Intuitively, this inference follows from the knowledge that if I were talking about Paul, I would be saying something trivially false. However, the inference that I was not talking about Paul cannot be made when the hearers would only look at the diagonal expressed by the utterance and the ambiguous pointing. To disambiguate, we need to look at the possible horizontal propositions expressed, where the horizontal propositions expressed are determined with respect to the possibilities of the context.

Determining the whole propositional concept is also needed for another case of ambiguity; the case of *questions*. According to the two-dimensional analyses of questions of Groenendijk & Stokhof (1982), and Lewis (1982), a question denotes its set of true answers. What a true answer is depends, of course, on how the world looks like. Thus, if A and B are the two relative alternatives, the *whether* phrase *whether A or B*, denotes a function from worlds to the proposition expressed by the disjunction of the true alternatives in this world. We can say that for any *w*, the *content* of *whether A or B* in *w*, $[A \veebar B](w)$, is defined as follows: $\{w' \in KI(w \models A \ \& \ w' \models A) \text{ or } (w \models B \ \& \ w' \models B)\}$. Thus, if in world *w* only A is the case, $[A \veebar B](w)$ equals $[A](w)$, and if in *w* only B is the case $[A \veebar B](w)$ equals $[B](w)$. Let us assume that a context, C, can be represented by a set of worlds. Because some worlds in context C might be A-worlds, and others B-worlds, $A \veebar B$ denotes a set of alternatives with respect to C: $\{[A \veebar B](w) | w \in C\}$, the set of possible true answers to the question. But it would be impossible to get such a set of alternative propositions if we didn't separate the roles of context and index; the diagonal proposition expressed by *whether A or B* is just the ordinary disjunction *A or B*. If we would not look at what horizontal proposition is expressed in each world, but only at the diagonal proposition, the question *Will John come?* states in each world the same trivial proposition: *John will come, or he won't come.*

By separating the roles of reference-context and index, we can also account for the two different uses of definite descriptions Donnellan (1966) pointed to. Consider the case where speaker and hearer see a woman with a man. The man treats the woman kindly, and speaker and hearer both assume, and know each other to assume, that the man is the woman's husband. However, the assumption is wrong; the man is not the woman's husband. As Donnellan argued convincingly, even if their assumption was wrong, we still have a substantial intuition that there is a sense in which the speaker said something true of the person he had in mind by uttering *Her husband is kind to her.*[94] Donnellan (1976) proposes to account for this fact by assuming that definite descriptions can be used in two ways: attributively, and referentially. If a description is used *attributively*, a speaker "states something about whoever or whatever is the so- and-so", in a *referential* use, a speaker "uses the description to enable his audience to pick out whom or what he is talking about and states something about that person or thing" (Donnellan, p. 285). It has always been unclear what kind of ambiguity Donnellan was pointing to, but it is generally agreed that we should not think of it as a *semantic* ambiguity.[95] But, then, how can we account for the

---

[94] Of course, when a third person informs the speaker and hearer that they are wrong about this guy, that he is not her husband, the speaker can normally no longer use the description *her husband* to refer to this man, and if so, only in a special 'ironical' sense.

[95] See Kripke (1977).

two readings of the above sentence if it is assumed that definite descriptions are *semantically* unambiguous; if they refer always to the unique (most salient) individual that satisfies their description in the world of consideration? Stalnaker (1970b) proposed a straightforward answer to this question; the referential/attributive ambiguity is not semantic, but *pragmatic* in nature. The referent is either determined by context, or depends on the relevant index and is determined by the proposition expressed. If a description is attributively used, the rule for determining the referent of the description is part of the horizontal proposition expressed. If in the actual world the man seen with the woman is not her husband, the description *her husband* does not refer to this man in the actual world, if the description is attributively used, but instead to the unique man, if any, who actually is her husband. If the description is referentially used, however, the rule for determining the referent of the description is not part of the horizontal proposition expressed, rather it will refer to the individual that is *presupposed* to be the unique (most salient) object that satisfies the descriptive content. If what is presupposed about the denotation of the description *her husband* is the same as what actually is the denotation of the description, it doesn't matter much whether the description is attributively or referentially used. However, there might be a difference if the actual world is not compatible with what is presupposed. If the speaker presupposes of *a* that he is her husband, the horizontal proposition expressed by the sentence *Her husband is kind to her* in which the description is referentially used will be that *a* is kind to her. This proposition might be true in the actual world, although *a* is not her husband. Thus, by separating the roles of context and index we can account for the intuition that in the above sketched circumstances *Her husband is kind to her* might have both a false, and a true reading, although the noun phrase is semantically unambiguous; in both cases the noun phrase refers to the unique (most salient) object that satisfies the description. Again, it is not possible to account for this intuition if one always looks at only the diagonal proposition expressed by a sentence.

The main claim of this section is that we can account for the phenomena dynamic semantics can account for if we assume that anaphoric pronouns are usually referentially used, because the pronoun picks up what is presupposed to be the relevant speaker's referent. Of course, if a pronoun or description is anaphorically used, there need not be a unique person that is presupposed to satisfy the descriptive content of the noun phrase, it is only presupposed that there is a (unique most salient) individual available to refer to by this pronoun or description because it satisfies its descriptive content. The speaker presupposes that there is such an individual available for reference by a pronoun, because he has acted in such a way that he has introduced a speaker's referent to the discourse. This can be illustrated by the contrast in acceptability between *John owns a donkey. Mary beats it* and **John is a donkey-owner. Mary beats it.* In contrast to the latter example, the indefinite *a donkey* is used in the former example which has a speaker's referent. In normal cases, the relevant act by which the speaker's referent is introduced is the speaker's use of an indefinite.

In an influential discussion of the proposal to treat pronouns as referential expressions, Heim (1982, § 1.3) argues exactly on the basis of the asymmetry in acceptability of the above two kinds of sentences against such a treatment of pronouns. She argues that this asymmetry cannot be predicted on the basis of the truth-conditions of the first sentences and the surrounding circumstances alone, because what seems crucial is how the utterance is worded. Heim observes that if pronouns are treated as variables bound by 'text-scope' existential quantifiers associated with *explicitly mentioned* indefinites, the asymmetry can be explained, and in later chapters she argues that this latter approach is indeed the way to go. But as Stalnaker (1996) rightly points out, this argument ignores the fact that a context contains not only information about the subject matter of conversation, but also the information which sentences are uttered in the conversation, and more general information about linguistic and speech conventions. As a result, also when pronouns are thought of as referential expressions the above asymmetry can be explained. The reason is that in the first sentence of the first example, but not in the second example, an indefinite is explicitly used, and it is a speech convention that only in the former case a pronoun can be appropriately

used, because normally when an indefinite is explicitly used, the speaker has a specific individual in mind.

Even if the speaker knows what individual is responsible for his use of the indefinite, the hearer need not. It can be that the only information the hearer has about this individual is that the speaker has intended to refer to it by his use of the indefinite. In general, it seems that a discourse referent stands for an individual with the only information associated with it that it verifies the sentence in which the indefinite occurs, and that the speaker intended to refer to it by his use of a certain indefinite. Thus, in general, the information the hearer has about a speaker's referent, and what the speaker can presuppose about it, can be thought of as the diagonalised speaker's referent. This, I wish to propose, is the information associated with a subject, or discourse referent

The speaker's referent of an indefinite will be relevant for the interpretation of pronouns that have an indefinite as their syntactic antecedent. Thus, if <<a,b>,w> is a reference-context/index pair in the context, as far as the hearer can tell, *w* might be the actual world in which the speaker has used two expressions which have *a* and *b*, respectively, as the speaker's referent. If the speaker uses the indefinite in *An S is P* specifically, he thereby enriches each possibility of the context with an object to which he, as far as the hearer can tell, could have referred. By the use of an indefinite, the speaker *changes* the reference-contexts of context C. If it is speaker's referent that count, it follows that it is impossible for C to contains two pairs <c,w> and <c',w> with the same index, but different reference-contexts, although there might be two such pairs <c,w> and <c,w'> in C with different indices but the same reference-context. The reason is that the speaker's referent is world-dependent. As a result, subjects can be thought of as individual concepts, functions from worlds to individuals.

If it is assumed that a context can never contain two pairs <c,w> and <c',w> with the same index, but different reference-contexts, it can be easily explained why a subject normally represents the *guise* under which hearers represent an actual speaker's referent. In the two-dimensional theory we are working in, the context C represents that what is common information among the participants of the conversation. As Stalnaker has stressed in various papers, this context contains not only information about the subject matter of conversation, but also information about the conversation itself. A context contains the first kind of information because all the assertions accepted by the participants of a conversation should be true in all the possibilities of the context. A context contains the second kind of information because it can be assumed that it is common information among the participants that the conversation is going on. Because each possibility of the context might be the actual world as far as the hearer can tell, in each of the worlds of the context (a counterpart of) the conversation is going on. In cases where the speaker has an individual in mind by his use of the indefinite, he introduces a speaker's referent in the actual world. Because it is unknown to the participants of the conversation what the actual world is, as far as they can tell, any world of the context might be the actual world in which the speaker introduced a speaker's reference by his use of (a counterpart of) the indefinite *in that world*. As a result, the guise under which the actual speaker's referent, if any, is represented to the hearers can be represented by an individual concept.

Although specific indefinites come with a speaker's referent, I follow Kripke (1977), Lewis (1979b) and Stalnaker (1996) in taking the speaker's referent of the indefinite to be semantically irrelevant to the interpretation of the indefinite itself. For the *truth* of a sentence of the form *An S is P*, only the *existential* information counts. That is, just like Kripke, Lewis and proponents of the dynamic theory stated in § 2.3, I disagree with Chastain (1975), Donnellan (1978) and Fodor & Sag (1982) who claimed that a sentence like *An S is P* is false if the indefinite is specifically used and does not refer to a *P*, although there is another object that is an *S* that is *P*. But that doesn't mean that it is semantically irrelevant what the object is the speaker had in mind for his use of the indefinite, as is assumed by DRT/FCS. The speaker's referent is relevant to semantics through pronominalisation.

Consider a case where I say *An S is P. It is Q* and have *a* in mind by the use of the indefinite. Suppose now that in fact *a* is an *S* that is *P*, but no *Q*, and that there actually is another object, *b*, that is an *S* that is *P* and *Q*. What would we say about the truth-value of the second sentence used in the above discourse? Geach, followed by proponents of standard CCT, would say that both sentences are true. I will follow Kripke (1977), Lewis (1979b), and Stalnaker (1996), however, in assuming that in these circumstances the first sentence is true, but the second false. By making this assumption, it is clear that the equivalence that holds in standard CCT between *An S is P. It is Q* and *An S that is P is Q*, represented in standard CCT by $\exists x[Sx \wedge Px] \wedge Qx$ and $\exists x[Sx \wedge Px \wedge Qx]$, respectively, has to be given up.[96]

Diagonalisation occurs only when a new object becomes available for a pronoun or (other) short definite description to refer to. This is typically the case when the speaker uses an indefinite, or when he points to an object. When the speaker uses a singular pronoun to refer back to such an object, no extra referential ambiguity will arise. Diagonalisation also takes place when a pronoun or definite description is demonstratively used, because in that case the use of the expression is accompanied by a pointing.

No two different reference-contexts can go with the same index, because the speaker's referent of the indefinite *An S* in *An S is P* will be the unique object causally 'responsible' for the agent's use of the indefinite. The diagonal proposition expressed by a sentence in which a pronoun occurs that picks up this speaker's reference is object-independent, not because in each world an existential proposition is communicated, but because in different worlds it could have been a different object that was responsible for the speakers use of the antecedent indefinite. According to this framework, identity conditions for worlds are stronger than in traditional modal logic; not only facts about the subject matter of conversation, but also facts about the acquaintance relations and about the discourse itself are relevant. If two worlds are identical with respect to subject matter, but differ with respect to what object was responsible for the speaker's use of the indefinite, two different referents would be introduced by the utterance of *An S is P* in the two worlds.

According to this view, there is an essential difference between what the speaker knows and what the hearer knows about the referent of a pronoun. In contrast to the hearer, the speaker normally knows what object he intends to talk about. It is clear that speaker *a* who has just said *A man is walking in the park* can continue his discourse with the sentence *He is walking his dog*. The speaker can refer to the same object with the pronoun *he* as he intended to refer to with the indefinite *a man*. But as noted by Groenendijk *et al.* (1995a), if it is presupposed that there are more men walking in the park, a different speaker, *b*, cannot continue the discourse by uttering the second sentence, if he doesn't know which of those men the speaker did intend to talk about. *B* can react, however, by saying *Then he is walking his dog*. By using the word *then*, *b* makes explicit that he doesn't know the speaker's referent, and intends to refer to *whoever a* had in mind by *a*'s use of the indefinite. *B*'s reaction is based on the conviction that *every* man walking in the park is walking his dog.

The different roles speakers and hearers play in dialogues with respect to the interpretation of indefinites can also explain how we normally interpret pronouns whose antecedents are indefinites in the antecedents of conditionals. In a donkey sentence like *If John owns a donkey, he beats it*, the indefinite *a donkey* does not have a speaker's referent. That is, it is not the *actual* reference-context that determines the referent of *it* in the consequent. Instead, we consider *hypothetical* or *counterfactual* reference-contexts (or worlds) in which the

---

[96] As we will see in § 2.5, Dekker (1994) and Groenendijk *et al.* (1994) also have given up this equivalence, but remarkably enough for an almost opposite reason.

indefinite had a speaker's referent.[97] These counterfactual reference-contexts should be as close to the actual reference-context as possible. In particular, if the antecedent holds in the actual world, the world of the counterfactual reference-context should make the same facts true about the conversation's subject matter as the actual world does.[98] Because the indefinite has no actual speaker's referent, the speaker has to be cautious. The consequent has to be true with respect to all such counterfactual reference-contexts.

By a Gricean maxim, in using a pronoun, the speaker must know the speaker's referent of its antecedent. Normally, speakers can only do this if they themselves are responsible for the introduction of this referent. Thus, a pronoun used in the 'main' context normally cannot pick up the referent of an indefinite used in the antecedent of a conditional.

The theory sketched in this section is very close to standard CCT as stated in § 2. But there are at least four important differences. The first difference is that for the above mentioned theories the status of discourse referents is not clear, while in the theory I have just sketched it is: a discourse referent or subject represents the information the hearer has about a certain speaker's referent, it is the diagonalised speaker's referent. The second difference comes out clearly in the truth definition. In both cases it is said that *A* is true in *w* with respect to context C if and only if there is a reference-context *c* such that <c,w> ∈ C and w∈ [A](c). But in contrast to the standard dynamic theories, in the theory I am proposing it is impossible for C to contain two pairs <c,w> and <c',w> with the same index but different reference-contexts, because in every world each use of an indefinite can have only one speaker's referent. It follows that I don't predict that the discourse *A man is walking in the park. He is whistling* will always be truth-conditionally equivalent to the sentence *A man who is walking in the park is whistling*, while this is predicted by the standard dynamic theories. The two are not equivalent when the speaker's referent of the indefinite is not whistling, although another man who is walking in the park is. This difference between the theory I favour and the standard dynamic theories will become relevant in the account of anaphoric dependencies across belief attributions, which will be discussed in chapter 3. The third difference is that my theory, by taking the notion of speaker's reference seriously, can offer an independently motivated explanation of why pronouns cannot always take an earlier mentioned indefinite as their antecedent. The pronoun *he* in the discourse *If a farmer owns a donkey, it is treated well. He is a nice person* normally cannot take the indefinite *a farmer* as antecedent, because this indefinite has no speaker's referent in the *actual* world. The fourth, and final, difference is that according to the approach sketched in this section it is clear what the limits are of this referential approach. I have argued that the referent of the pronoun is determined by the intention of the speaker. In case the pronoun is referentially used, what counts is the *specific intention*. But not all pronouns are referentially used: the referent of a pronoun does not always depend on the specific intention of the speaker. As I will argue later in this chapter, sometimes it is only the *general intention* that counts, and that the constraints on anaphora for these cases will be completely different. All these differences between the standard dynamic account and the approach sketched here are, I believe, happy consequences of the present account.

---

[97] Fiction and pretence, *There once was a beautiful princess*, also have to be analysed in terms of counterfactuallity, cf. Lewis (1978).

[98] I will assume that these facts must be true at the moment of utterance. Sometimes, however, a conditional says something about facts of the *future*, as in *If John marries a girl his parents disapprove of, they will make life quite unpleasant for her* (Partee, 1972). In these cases, I wish to propose, the pronoun *her* will be a descriptive pronoun as will be implemented in § 2.6. (See also chapter 6 for a similar analysis of attitude attributions conditionally dependent on desire ascriptions.) Paul Dekker (personal communication) suggested to me that something similar might be done to account for the weak reading of *If I have a dime in my pocket, I put it into the meter*, according to which the sentence is true if I put one dime into the meter, if I have at least one in my pocket. We consider hypothetical *future* worlds in which I am looking for a dime, and refer with *it* to the first one I will find.

According to the above suggestions, information states should contain more information than is assumed in standard dynamic semantics. Subjects should be thought of as the hearer's representation of the speaker's referent of a certain indefinite. To formally account for this extra information we can in principle go two ways: either we take the DRT-approach and complicate the representation of sentences in which indefinites occur, or we take the FCS-approach and complicate the possibilities that represent the information presupposed. In a DRT-like account we could represent a sentence like *An S is P*, which originally was represented by $\exists x[S(x) \wedge P(x)]$, now by $\exists x[S(x) \wedge R^S(s,x) \wedge P(x)]$. In the latter formula, '$R^S(s, x)$' is then supposed to mean something like 'the unique individual who was 'responsible' for the speaker's use of *An S*'. Although something like this might be possible, I think it is more straightforward to account for the problems discussed here by enriching the possibilities of the context. The major problem a representational account has to face in taking the notion of speaker's reference seriously is to account for donkey sentences in such a way that we still can give a uniform meaning to indefinites needed for a compositional analysis. The reason is obvious: in these cases the indefinite has no actual speaker's reference. I suggested earlier that in these cases we should look at *hypothetical* reference-contexts, or *counterfactual* worlds, where the indefinite had a speaker's reference. The easiest, although not the best, way to implement this suggestion is to determine speaker's referents by means of *choice functions*.[99] A choice function, ϕ, is a function from sets of individuals to individuals, such that the individual is an element of this set. The idea is that each speaker in each world has a certain perspective onto this world, and that this perspective is modelled by a choice function. If speaker *a* uses the indefinite *A man* in *w*, he intends to refer in *w* to the man chosen by his choice function in *w*, $\phi(I_w(\text{man}))$.[100] In every world, every speaker has only one perspective, or choice function. That is, if the only difference between *w* and *w'* is that the perspective of speaker *a* can be represented by choice function ϕ in *w*, and by ϕ' in *w'*, it is the case that *w* and *w'* differ from each other if ϕ and ϕ' differ from each other. Possibilities are no longer represented by world-assignment pairs, but by triples like $<\phi,g,w>$, where the choice function ϕ accounts for the introduction of the speaker's referents by the use of indefinites, and the partial assignment function *g* for the interpretation of pronouns. I will assume that indefinites are represented by epsilon terms, and that a sentence like *A man is walking in the park with her* is represented by something like '$\text{WiP}(\varepsilon x\text{Mx}, y)$'.[101] The interpretation of this formula with respect to context S will have the effect that (i) all worlds are eliminated in which there is no man who is walking with the referent of *y* in the possibility, (ii) the speaker's referent associated with the indefinite *a man* is introduced, and (iii) the

---

[99] See also Reinhard (1992), Kratzer (ms.) and Winter (1996), although in contrast to what I argued for, the two former authors assume that the speaker's referent is semantically relevant for determining the truth-conditions of the sentence in which the indefinite occurs. They concentrated on the phenomenon that sometimes a pronoun used in the 'main' context can refer back to an indefinite in the antecedent clause of a conditional, as in *If a plumber comes, let him in. He will be coming to repair the bathtub*. It seems that the pronoun *He* refers back to the actual speaker's referent of the indefinite *a plumber* in the antecedent of the conditional. According to recent syntactic theories, antecedents of conditionals are so-called scope islands; quantifiers cannot outscope conditionals. Reinhard (1992) and Winter (1996) argue that to account for the phenomenon we should assume that it is not the indefinite *a plumber* that takes scope over the conditional, but just the choice function, instead. I would prefer to follow Kratzer (ms.) in accounting for this phenomenon, if possible, in terms of double indexing and the *dthat* operator, and not in terms of scope. I won't try to make this suggestion precise, however.

[100] In fact, it is not really the set of men in the world of consideration that counts. The phenomenon of pronominal contradiction clearly shows this. The speaker's referent is not dependent on the *type* of the indefinite used, but on the *token*. The idea that speaker's reference depends on the *beliefs* and *intentions* of the speaker related with a certain use of an indefinite should be taken more seriously than I do here.

[101] One might think that by treating indefinites as terms, the scope differences of indefinites can no longer be accounted for. But that is not true, because we can account for scope differences by using the abstraction operator as in chapter 1. See § 2. 8 for an interpretation of the abstraction operator in dynamic semantics.

choice function changes in such a way that a next use of the indefinite *a man* will not have the same individual as its speaker's referent:

$$[[\text{WiP}(\varepsilon x \text{Mx}, \text{y})]](\text{S}) = \{\langle\phi',\text{h},\text{w}\rangle \mid \exists \text{g}, \phi: \langle\phi,\text{g},\text{w}\rangle \in \text{S} \;\&\; \langle\phi(\text{I}_\text{W}(\text{M})), \text{g}(\text{y})\rangle \in \text{I}_\text{W}(\text{WiP}) \;\&\; \text{g}[\text{x}]\text{h} \;\&\; \text{h}(\text{x}) = \phi(\text{I}_\text{W}(\text{M})) \;\&\; \phi' = \phi \text{ except that } \phi'(\text{I}_\text{W}(\text{M})) = \phi(\text{I}_\text{W}(\text{M}) - \{\phi(\text{I}_\text{W}(\text{M}))\})\}$$

I will say that a sentence A is true in possibility $\langle\phi,g,w\rangle$ if and only if there is a choice function ϕ' such that $[[A]](\{\langle\phi',g,w\rangle\}) \neq \varnothing$. Although for truth of the sentence, the actual choice function is irrelevant, the individual that is introduced and can be referred back to later by a pronoun *does* depend on this actual choice function. Thus, in possibility $\langle\phi,g,w\rangle$ the man that is referred to and thus introduced is $\phi(\text{I}_\text{W}(\text{Man}))$, and the possibility is changed into $\langle\phi',h,w\rangle$ such that $\phi'(\text{I}_\text{W}(\text{Man})) \neq \phi(\text{I}_\text{W}(\text{Man}))$ (to account for novelty), and *h* is the same as *g* except that *x* is also in the domain of *h* and $\text{h}(\text{x}) = \phi(\text{I}_\text{W}(\text{Man}))$. A following sentence like *he whistles* can now be represented by 'W(x)' and interpreted like in standard CCT as follows:

$$[[\text{W}(\text{x})]](\text{S}) = \{\langle\phi,\text{h},\text{w}\rangle \in \text{S} \mid \text{h}(\text{x}) \in \text{I}_\text{W}(\text{W})\}$$

The nice thing about using choice functions to determine speaker's reference is that in this way we can give a uniform analysis of indefinites. The reason is that indefinites under the scope of a negation can still be interpreted with respect to choice functions, but in these cases it is not the choice function that represents the speaker's perspective that counts for the introduction of the referent, but any arbitrary choice function instead. Like in Hintikka's Game Theoretic Semantics,[102] it's not the speaker that is responsible for the reference of indefinites under the scope of a negation, but the hearer, or nature. The interpretation rule for negation looks as follows:

$$[[\sim\text{A}]](\text{S}) = \{\langle\phi,\text{g},\text{w}\rangle \in \text{S} \mid \sim\exists\phi', \text{h}: \text{g} \subseteq \text{h} \;\&\; \langle\phi',\text{h},\text{w}\rangle \in [[\text{A}]](\{\langle\phi'',\text{k},\text{w}'\rangle \mid \exists\phi':\langle\phi',\text{k},\text{w}'\rangle \in \text{S}\})\}$$[103]

Negation is simply interpreted as a perspective shifter. It follows that if a donkey sentence like *If a farmer owns a donkey, he beats it* is represented by something like $\text{Own}(\varepsilon x\text{Fx}, \varepsilon y\text{Dy}) \rightarrow \text{Beat}(x,y)$, it is true in *w* iff for every farmer-donkey pair that stands in the own-relation in *w*, it holds that the farmer of this pair beats the corresponding donkey, just like in original CCT.

Until now I have only looked at atomic clauses and negation, and for atomic clauses I only looked at formulae that introduce only one object and only at terms corresponding with indefinites without relative clauses. Let me now quickly generalise the above picture, and make things a bit more precise. First, I give the new interpretation rules for formulae:

$$(1')\quad [[\text{R}(t_1,\ldots,t_n)]](\text{S}) = \{\langle\phi',\text{h},\text{w}\rangle \mid \exists\text{g}, \phi:\langle\phi,\text{g},\text{w}\rangle \in \text{S} \;\&\; \text{h} = \text{g} \cup \text{Term}(\text{R}(t_1,\ldots,t_n),\langle\phi,\text{g},\text{w}\rangle) \;\&\; \langle[[t_1]]^{\phi,\text{h},\text{w}},\ldots,[[t_n]]^{\phi,\text{h},\text{w}}\rangle \in \text{I}_\text{W}(\text{R})$$

---

[102] See for instance Hintikka & Kulas (1985).

[103] Note that by proposing this interpretation rule, I make the assumption that some contexts might contain two pairs like $\langle c,w\rangle$ and $\langle c',w\rangle$ with the same index, but a different reference-context. I have argued that this should never be the case, but I can only remedy this in § 2. 10 where I account for CCT in terms of possible worlds only.

& $\phi' = \phi$ except that for all $x \in$ VAR and all i: $1 \leq i \leq n$: if $t_i = \varepsilon x P$, then $\phi'(I_W(P)) = \phi(I_W(P) - \{\phi(I_W(P))\})\}$

(2') $[[\sim A]](S) = \{<\phi,g,w> \in S \mid \sim\exists\phi', h: g \subseteq h$ & $<\phi',h,w> \in [[A]](\{<\phi'',k,w'> \mid \exists\phi': <\phi',k,w'> \in S\})\}$

(3') $[[A \wedge B]](S) = [[B]]([[A]](S))$

(5') $[[\forall xA]](S) = \{<\phi,g,w> \in S \mid \forall h[g[x]h \Rightarrow \exists\alpha[<\phi,h,w> \leq \alpha$ & $\alpha \in [[A]](S[x])]]\}$

The two possibilities $<\phi,g,w>$ and $<\phi',h,w'>$ stand in the order relation $\leq$, $<\phi,g,w> \leq <\phi',h,w'>$, iff $g \subseteq h$ and $w = w'$. The notation 'S[x]' is an abbreviation for the set $\{<\phi',h,w> \mid \exists\phi, g: <\phi,g,w> \in S$ & $g[x]h\}$.

The above interpretation rules depend on the following definitions of the predicate *Term*, and $[[A]]^{\phi,g,w} = 1$:

Term(t, $<\phi,g,w>$) $= \{<x,d>\} \cup$ Term(A, $<\phi,g[^x/_d],w>$), if $t = \varepsilon_x A$ & $\phi(\{d' \in D \mid [[A]]^{\phi,g[x/d'],w} = 1\}) = d$,
$= \emptyset$, if $t \in$ Var

Term($R(t_1,...,t_n)$, $<\phi,g,w>$) = Term(t_n, $<\phi,g \cup$ Term(t_{n-1}, $<\phi,g \cup$ Term(...Term(t_2, $<\phi,g \cup$ Term(t_1, $<\phi,g,w>$),w>)...),w>)>)

Term($\sim A$, $<\phi,g,w>$) $= \emptyset$

Term($A \wedge B$, $<\phi,g,w>$) = Term(B,$<\phi,g \cup$ Term(A, $<\phi,g,w>$),w>)

Term($\forall xA$, $<\phi,g,w>$) $= \emptyset$

$[[R(t_1,...,t_n)]]^{\phi,g,w} = 1$ iff $h = g \cup$ Term($R(t_1,...,t_n)$, $<\phi,g,w>$) & $<[[t_1]]^{\phi,h,w},...,[[t_1]]^{\phi,h,w}> \in I_W(R)$

$[[\sim A]]^{\phi,g,w} = 1$ iff $[[A]]^{\phi,g,w} \neq 1$

$[[A \wedge B]]^{\phi,g,w} = 1$ iff $[[A]]^{\phi,g,w} = 1$ and $[[B]]^{\phi,h,w} = 1$, where $h = g \cup$ Term(A, $<\phi,g,w>$)

$[[\forall xA]]^{\phi,g,w} = 1$ iff for all $d \in D(w)$, there is a ϕ': $[[A]]^{\phi',g[x/d],w} = 1$

Now I can define the most important semantic concepts. A formula A is *acceptable in S*, $S \models A$, if and only if S is a substate of [[A]] (S), in the sense that every $\alpha \in S$ can be extended to an $\alpha' \in [[A]](S)$ such that $\alpha \leq \alpha'$. The discourse $A_1 \ldots A_n$ is *true in* $<\phi,g,w>$, $<\phi,g,w> \models A_1 \ldots A_n$, if and only if there is a ϕ' such that $\{<\phi',g,w>\} \models A_1$ and $<\phi,g \cup$ Term(A_1, $<\phi,g,w>$),w$> \models A_2 \ldots A_n$.[104] A *classically entails* B *with respect to* g, $A \models_g B$, if and only if $\{w \in W \mid \exists\phi: <\phi,g,w> \models A\} \subseteq \{w \in W \mid \exists\phi: <\phi,g,w> \models B\}$. A *diagonally entails* B, $A \models_d B$, if and only if for all S: $[[A]](S) \models B$.[105]

In the theory I have just formulated, information states contain more information than the information states in standard CCT. As a result, the semantic concepts of acceptability, truth, and entailment defined in § 2.3 can also be defined in terms of the information

[104] I thank Maria Aloni for spotting a mistake in my truth-definition of an earlier version.

[105] Compare the definitions of classical and diagonal entailment with the ones given in § 1.8.

available in the present theory. That is, we need only one extra concept: an equivalence relation between worlds with respect to the subject matter of conversation. Let us say that 'SSM' denotes this relation. In that case, we can say that the discourse $A_1 ... A_n$ is true in $<\phi,g,w>$ if and only if there is a ϕ' and a w' such that SSM(w,w') and $\{<\phi',g,w'>\} \models A_1 \wedge ... \wedge A_n$. As a result, I am not claiming that standard CCT says anything wrong, I'm just claiming that it doesn't say enough; it should take the notion of speaker's reference more seriously than it actually does. In the next chapter I will argue that the notion of *speaker's reference* is also relevant to semantics in order to give an appropriate analysis of intentional identity attributions.

In this section I have argued that pronouns are normally referentially used, and showed that this view can be formalised, too. For convenience, however, I will use in the rest of almost this whole chapter the formulation of standard CCT as given in § 2.3. Only after I have argued that we could account for anaphoric dependencies in terms of salience in § 2.9, in the final section of this chapter I will make use of the proposal made here to account for CCT in terms of possible worlds only.

2.5 Epistemic might

Just as in the traditional account of Stalnaker (1978), for metaphysical necessity, we check whether the horizontal proposition is necessary true, for epistemic possibility, on the other hand, we check whether the diagonal proposition is consistent with the context. Indeed, Veltman (1990) interpreted epistemic *might* in his update semantics, which is limited to propositional logic, as follows (if *I* is a set of worlds):

$$[[\Diamond A]](I) = I, \text{ if } [[A]](I) \neq \varnothing,$$
$$= \varnothing \text{ otherwise.}$$

Recently, it has been an issue how we should account for the integration of Veltman's (1990) epistemic *might* into CCT.[106] That there can be an issue is due to the fact that updates in original CCT (without definedness conditions) and Veltman's update semantics each have a property that the other system lacks. The relevant properties are *distributivity* and *eliminativity*:

An update [[A]] is *distributive* iff $\forall S$: $[[A]](S) = \cup_{\alpha \in S}[[A]](\{\alpha\})$

An update [[A]] is *eliminative* iff $\forall S$: $[[A]](S) \subseteq S$.

Veltman's Update Semantics is *eliminative*, but because of the *might* operator *non-distributive*,[107] while original CCT is *distributive* but *not eliminative*. As in all dynamic systems, updates in both frameworks are non-commutative. The greatest common devisor is distributive and eliminative, and thus not dynamic, that is, non-commutative any more.[108] The problem that arises with the integration of the two theories is how to combine quantification with epistemic modalities.

---

[106] There is more to epistemic *might* than I will discus in this section. For instance, we have to account for anaphoric dependencies that original CCT cannot account for as in *A thief might break in. She might steal the silver.* In § 2.7, I will account for those anaphoric dependencies, too.

[107] But see § 2.10.

[108] We have to prove that for all A, B and S:$[[A \wedge B]](S) = [[B \wedge A]](S)$. So, $\alpha \in [[A \wedge B]](S)$ iff $\alpha \in [[B]]([[A]](S))$ iff (by distributivity) $\alpha \in \cup_{\alpha' \in S}[[B]]([[A]](\alpha'))$ iff $\exists \alpha' \in S$ & $\exists \alpha''$: $\alpha'' \in [[A]](\{\alpha'\})$ and $\alpha \in [[B]](\{\alpha''\})$ iff (by eliminativity) $\alpha \in [[A]](\{\alpha\})$ and $\alpha \in [[B]](\{\alpha\})$ iff $\alpha \in \cup_{\alpha' \in S}[[A]]([[B]](\alpha'))$ iff $\alpha \in [[A]]([[B]](S))$ iff $\alpha \in [[B \wedge A]](S)$.

Just like propositional update semantics must account for the fact that "It might be that A. ... It is not the case the A" is consistent, but "It is not the case that A. ... It might be that A" is not, the system resulting from the integration of epistemic *might* into CCT must account for the fact that "$\exists$xPx $\wedge$ $\Diamond$Qx $\wedge$ ~Qx" is consistent, but "$\exists$xPx $\wedge$ ~Qx $\wedge$ $\Diamond$Qx" is not.

The simplest way to integrate the epistemic modalities into a formulation of CCT that has the update property is simply to leave everything as it is, so, use Veltman's clause for the interpretation of *might* with *I* replaced by *S*, and also leave the interpretation of the quantifiers as given above. Dekker (1993) actually did this. However this proposal leads us into *Dekker's Problem*, as he observed himself. $\exists$x$\Diamond$Ex outputs all possible values of *x* if some value of *x* is possibly an E. It should however only output those values of *x* that are possibly an E. In the same way, $\forall$x$\Diamond$Ex is satisfied in any state where there is a possibility of somebody having property E, even if there are some individuals of whom it is known that they don't have property E. So, although the system correctly predicts that "$\exists$y~Ey & $\forall$x$\Diamond$Ex" is consistent, it also predicts that "~Ea & $\forall$x$\Diamond$Ex" is consistent, even if *a* is a constant that denotes a particular object, the same in every possibility. The problem arises because we consider all the values of *x* simultaneously, and there is no way of checking that Ex is consistent with some choices of a value, but not with others.

But Dekker's proposal does not only give rise to Dekker's problem, there is also a *donkey-closet problem* to be solved. In CCT as I have given it in § 2.3, the formulae $\exists$xA $\wedge$ B and $\exists$x[A $\wedge$ B] are equivalent, and the same holds for $\exists$xA $\rightarrow$ B and $\forall$x[A $\rightarrow$ B]. For obvious reasons, the equivalences in CCT are known as the *donkey equivalences*. However, once epistemic *might* is introduced into the language, the equivalences don't hold anymore. Consider the following sentences:

(7) There is someone hiding in the closet. He might be the one who did it.
(8) There is someone hiding in the closet who might be the one who did it.
(9) If there is someone hiding in the closet, he might be the one who did it.
(10) Anyone who is hiding in the closet might be the one who did it.[109]

If the donkey equivalences would hold for all kind of constructions, (7) would be equivalent to (8), and the same should hold for (9) and (10). However, given the following model:

$W = \{w, w'\}$, $D = \{d,d'\}$, $I_W(R) = \{d\}$, $I_{W'}(R) = \{d'\}$, $I_W(Q) = I_{W'}(Q) = \{d\}$

and given that we presuppose that either *w* or *w'* is the actual world, intuitively after interpreting (8), $\exists$x[Rx $\wedge$ $\Diamond$ Qx], we like to end up in the information state that we are in *w*, that is that *d* is *R*. After updating our information state with (7), $\exists$xRx $\wedge$ $\Diamond$ Qx, however, we still don't know in what world we are, thus who the *R* is. The same holds for respectively (10), $\forall$x[Rx$\rightarrow$ $\Diamond$ Qx], and (9) $\exists$xRx $\rightarrow$ $\Diamond$ Qx.[110]

Dekker (1994) and Groenendijk *et al.* (1994) propose to give up the above donkey-equivalences. They give up these equivalences by proposing an alternative interpretation rule for indefinites:

$[[\exists xA]](S) = \cup_{d \in D}[[A]](S[x:= d])$, if $\forall g \in G(S)$: $x \notin dom(g)$,

[109] Groenendijk *et al.* (1994) attribute the sentences to David Beaver.
[110] At least, these are the facts according to Groenendijk *et al.* (1994).

undefined otherwise[111]

By this new interpretation rule for indefinites a semantic distinction is made between what Evans would call *bound* and *unbound* pronouns. Bound pronouns are interpreted rigidly, as real individuals, whereas unbound pronouns are not. By interpreting bound pronouns as rigid individuals, *Dekker's problem* is solved. The formula $\exists x \Diamond Ex$ only output those values of x that are possibly an E, and $\forall x \Diamond Ex$ will be satisfied in a state just in case there is no individual in the domain of quantification whose being E is impossible. By making a distinction between bound and unbound pronouns also the *donkey-closet problem* is solved. Sentences (8) and (10) are more informative than (7) and (9) because contrary to the latter sentences, in the former ones we consider rigid individuals individually when epistemic *might* is interpreted.

By the way information states in CCT are modelled, it follows that epistemic *might* quantifies over more fine grained entities than possible worlds insofar as they represent only the information about the subject matter of the discourse.

Still, there are some worries. Would the speaker of a sentence like $\Diamond A$ really ever express his incomplete knowledge about more than the subject matter of the discourse? And does $\exists xPx \rightarrow \Diamond Qx$ really have even weaker truth conditions than $\exists xPx \wedge Qx$, and thus say something different from $\forall x[Px \rightarrow \Diamond Qx]$, when it is not presupposed that there is at most one P? I am not convinced, but I can only give more content to these worries after I have argued that also CCT is in need of

2.6 Descriptive pronouns[112]

One of the main motivations in § 2.4 for treating at least some uses of pronouns as being referential was the phenomenon of *pronominal contradiction*. A pronoun sometimes refers back to the speaker's referent of the antecedent. But as Kripke (1977) noted, the same phenomenon shows that sometimes pronouns also refer back to the *semantic referent* of the antecedent. If A says *Her husband is kind to her*, B can react by saying *No, he isn't. The man you are referring to isn't her husband.* The pronoun *he* refers to the semantic referent of its antecedent, the actual husband of her. Thus, pronouns can be used in two ways: they either pick up a previous semantic reference or a previous speaker's reference.

In § 2.4, I argued that pronouns are normally referentially used, referring to the unique individual that is the speaker's referent of the antecedent of the pronoun. According to the motivation I gave for CCT it can be explained why an indefinite used under the scope of two negations, can normally not be taken as syntactic antecedent for pronouns used in the 'main' context. The reason is that the indefinites occurring in those positions normally have no actual speaker's referent, something that is required if the pronoun is referentially used. In a similar way it can also be explained why an indefinite used in one disjunct can normally not be taken up by a pronoun occurring in the other disjunct. In original CCT as given in § 2.3, such constraints are given by syntactic means. Unfortunately, there are well known counterexamples to these constraints on anaphoric binding; pronouns can sometimes take an indefinites as syntactic antecedent, although the anaphoric island constraints predicted by DRT/FCS are violated. It has been observed that these cases can be

---

[111] Beaver (1993) showed that once presuppositions are considered Dekker's problem has its Heimean variant. Assuming the satisfaction account and the interpretation rule for indefinites as in original CCT, Heim (1983) predicted that *A man loves his cat* presupposes that *Every man has a cat*. Beaver suggested that by using the interpretation rule for indefinites proposed by Groenendijk *et. al.*, quantified sentences get much more reasonable presuppositions. And indeed, for indefinites occurring in non-distributive positions he seems to predict just right.

[112] Although the case for the existence of E-type pronouns is usually found most convincing with respect to plural pronouns, I will limit myself in this section to singular pronouns.

accounted for by assuming that these pronouns should be interpreted as E-type pronouns. Indeed, that's what I want to argue for, too. But maybe contrary to other proposals, I claim that there is a natural motivation for this division of labour, and I formally account for the existence of E-type pronouns in CCT. The motivation for the division of labour I gave above: What a speaker refers to with his use of a pronoun depends on his intentions, normally pronouns are referentially used, and it are the specific intentions that count, but sometimes pronouns are descriptively used, because it are only the general intentions that count. In the latter cases, a pronoun refers to the unique individual, if there is any, that satisfies the description associated with the pronoun that is recoverable from the antecedent clause. Note that if it is assumed that pronouns can be descriptively used, there is no reason anymore to expect that pronouns cannot escape the DRT/FCS anaphoric island constraints. Instead, the constraints for the appropriate use of descriptive pronouns should be given in terms of what is presupposed to be true.[113]

In this section I will concentrate only on one kind of example where the constraints on anaphoric binding of CCT as given above are too rigid. I will focus my discussion on a recent paper of Krahmer & Muskens (1995) where it is implicitly claimed that we don't have to rely on the existence of descriptive pronouns to account for some apparent counterexamples of CCT.[114] I will argue, however, that their proposal leaves something unexplained, and that this can be accounted for naturally by the E-type approach. Later, I will account for the existence of descriptive pronouns in CCT.

Consider the following sentences:

(11) Either John does not own a donkey or he keeps *it* very quiet. (Evans, 1977)
(12) Either there is no bathroom in the house, or *it*'s in a funny place. (Roberts, 1989).[115]
(13) It is not true that John didn't bring an umbrella. *It* was purple and *it* stood in the hallway. (Muskens & Krahmer, 1995)

It is well known that the standard CCT has problems with such sentences. The reason is that in CCT negation is treated as a plug with respect to anaphoric binding. Note that contrary to the standard dynamic approach, negations don't have this property according to the E-type account. Proponents of the standard dynamic account argue that negations *should* be treated as plugs, how else to account for the unacceptability of (14)?

(14) There is no guest at this wedding. *He* is standing right behind you.

The unacceptability of (14) can be accounted for by *syntactic* means. An object 'introduced' under the scope of a negation cannot be picked up by anaphoric means in further discourse. But the E-type approach has, of course, no problem with the unacceptability of (14). The sequence (14) is out, not for syntactic but for *semantic* reasons. The context resulting after the interpretation of the first sentence of (14) contains no world in which there is a guest at this wedding. If the pronoun *he* of the second sentence would stand for *the guest at this wedding*, the second sentence would be trivially false. That's the reason why (14) is out. That is quite a natural reasoning, I would say. And does the acceptability of the sentences (11), (12) and (13) not justify this reasoning?

Not so, say Krahmer & Muskens (1995). Negation is a syntactic plug with respect to anaphoric binding, and the reason why (11) - (13) are acceptable is that a double negation is a plug unplugged. A clause of the form ~~A is not only truth-conditionally, but also

---

[113] Maybe after accommodation.
[114] Somewhat similar proposals are made by Groenendijk & Stokhof (1990) and Dekker (1993).
[115] Attributed to Partee.

dynamically equivalent with A. They can account for this claimed equivalence in a way that is not completely ad hoc by using techniques from partial logic.[116]

There are some worries with their approach, however. First, intuitively there seems to be no difference between (11) and a sentence like:

(15) It is possible that John does not own a donkey,
but it is also possible that he keeps *it* very quiet.

It would be nice if both could be handled by the same mechanism. But it is rather doubtful that this mechanism will be that ~~A is equivalent to A. Second, if an indefinite is used under the scope of two negations, it seems that a singular pronoun can only take it as syntactic antecedent if there is only one object (in each of the relevant worlds) that could be the referent of the indefinite. For (11) and (12), for instance, the uses of the pronoun *it* in the second disjuncts can only pick up the *unique* donkey that John owns, and the *unique* bathroom in the house, respectively. If it is presupposed that possibly John owns more donkeys, and if there are maybe more bathrooms in the house, the uses of *it* in the respective second disjuncts would be, I think, inappropriate.

Moreover, my worry is not limited to disjunctions. I think that if an indefinite is used under the scope of two negations, a singular pronoun that is not standing under the scope of these negations can *never* take it as syntactic antecedent if there are more objects in one of the worlds that the indefinite could have referred to. It seems that Krahmer & Muskens agree. Discussing the contrast in acceptability between (16a) and (16b),

(16a) It is not true that there is no guest at this wedding.
*?He is standing right behind you.
(16b) It is not true that there is no bride at this wedding.
She is standing right behind you.

they say that the distinction is due to a uniqueness effect.

> Given some highly unlikely context in which it is understood between speaker and hearer that at most one guest can be present at this particular wedding (16a) would be fine. We feel that it is precisely the unlikelyhood of such a context which explains the markedness of (16a). (Krahmer & Muskens, 1995, p. 359)

I completely agree. But then they make the following claim about these problematic cases:

> Since such apparent counterexamples on closer examination turn out to be no counterexamples at all, it seems we can take it as a general rule that as far as truth conditions and the possibility of anaphora are concerned double negations in standard English behave as if no negation at all were present. (Krahmer & Muskens, 1995, p. 359)

I'm afraid that I don't understand this. That you can explain why a counterexample to your approach *is* a counterexample, doesn't mean that on closer examination it 'turns out to be no counterexample at all'.[117]

I want to propose to take the counterexample seriously. The speaker can appropriately use a singular pronoun that takes an indefinite as its syntactic antecedent although original CCT predicts that the antecedent is not accessible only in case it is presupposed, understood between speaker and hearer, that there is exactly one object that the indefinite can refer to.

---

[116] Groenendijk & Stokhof (1990) and Dekker (1993) reach a similar result in a less ad hoc way by using Amsterdam-lifting instead of Tilburg-partiality.

[117] It is sometimes assumed that we can account for bathroom sentences by representing sentences of the form "~P or Q" by something like "~P ∨ (P ∧ Q)". But this does not only give rise to the same problem as the approach of Krahmer & Muskens, it is also purely ad hoc.

This is also what's going on, I think, in Peirce's puzzle as discussed by Gillon (1996). In standard CCT, and in ordinary predicate logic, (a) $\exists xA \vee \exists yB$ is logically equivalent with (b) $\exists x(A \vee B)$. However, it seems that we have to translate a sentence like *Either someone wouldwin $1.000, if everyone took part, or someone will not take part* by $\exists x[A > Cx] \vee \exists yB$ and *Either someone will win $1.000, if everyone would take part, or he will not take part* in standard CCT by $\exists x[(A > Cx) \vee B]$ (to account for co-reference) that are not equivalent to each other. To see this, consider the following situation: There is a sweepstakes in which only one thousand people are eligible to participate. Tickets are sold for $1 each. No participant is permitted to buy more than one ticket. And the winner will take the total of the stakes, if everybody takes part. In this situation, the first sentence is true, but the second false. This example does not falsify the equivalence between $\exists xA \vee \exists yB$ and $\exists x(A \vee B)$ in ordinary predicate logic, of course. It only shows that we should not represent the second sentence, *Either someone will win $1.000, if everyone would take part, or he will not take part*, by a formula like '$\exists x[(A > Cx) \vee B]$', but by $\exists x[A > Cx] \vee B]$, instead. Indeed, Gillon observes that if the existential quantifier associated with *someone* is given smaller scope than the disjunction, and if the pronoun *he* is interpreted as a descriptive pronoun, we can account for the fact that the two sentences have a different meaning. The one who would win $1.000 should everyone take part, might still take part and win, but less than $1,000, because someone else does not take part.

I claimed that singular descriptive pronouns can only take an indefinite as syntactic antecedent if there is exactly one object that the description recoverable from the antecedent clause can refer to.[118] This picture should be slightly corrected, though. There is at least one singular pronoun that intuitively picks up more objects, although the pronoun itself *refers* only to one object. The pronoun I have in mind is *one*. As is commonly assumed, this singular pronoun takes a noun as syntactic antecedent.[119] What is relevant here is that this pronoun is a special kind of E-type pronoun in that it takes up properties and violates the CCT constraints on anaphoric binding in exactly the same way as other descriptive pronouns. Note, for instance, that the following variant of (16a) is perfectly o.k.:

(16a') It is not true that there is no guest at this wedding.
One is standing right behind you.

Below, I will give an interpretation rule for this quantified pronoun, but leave it to the reader how nouns introduce properties.[120]

It is tempting to propose that the division of labour between CCT and the E-type approach should be taken serious in the following way: Where CCT accounts for pronouns that take *specifically* used indefinites as antecedents (that is, either the speaker has a specific individual in mind, or the hearer is allowed to take any specific referent), E-type pronouns can only take indefinites as antecedent where *the speaker* is responsible for the referent of

---

[118] For extra motivation, consider the following contrast observed by Partee (1972) between *John was looking for the man who murdered Smith, and Bill was looking for him too*, and *John was looking for a gold watch, and Bill was looking for it too.* The pronoun *him* in the former sentence can be used when the speaker has no particular man in mind, but the pronoun *it* in the latter sentence cannot be used when the speaker has no particular watch in mind. The reason is, according to Partee, that in the latter case it cannot be presupposed that there is exactly one golden watch.

[119] Consider Partee's (1972) *John lost a black pen yesterday and Bill found a grey one today.*

[120] In fact, nouns do not just introduce properties, they introduce themselves as words, too. Any word that we have used has introduced itself, and we can refer back to it by anaphoric means with abstract pronouns: A: That's a rhinoceros. B: Spell *it* for me (Lyons, 1977). The context change theory should really be as fine grained as the token analysis Stalnaker (1981) suggested. I ignore this issue.

the indefinite, but had no specific individual in mind for his use of the indefinite. But I believe this is not the way to go. Again, the phenomenon of pronominal contradiction can be our guide. In Kripke's (1977) example, if A says *Her husband is kind to her*, he has a specific individual in mind, still a hearer might react by saying *No he isn't. The man you are referring to isn't her husband*, where the pronoun is descriptively used. So, at least if we also want to account for dialogues, I think we should not give separate interpretation rules for respectively specifically and unspecifically used indefinites.[121]

I conclude that to account for E-type pronouns in CCT, we have to implement the following ideas. First, you can always refer back with a singular pronoun to an indefinite if in each world of the relevant context there is only one object that could be a referent of the indefinite that verifies the antecedent sentence. For that reason, second, negations should not be treated as absolute syntactic plugs with respect to anaphoric binding.[122]

Although original CCT will be changed, the result will be closely related to it. Those cases that were solved by original CCT, especially donkey sentences, I treat in the same way. I only use instead of one, two assignment functions. The first one represents properties introduced by indefinites, while the second one assigns individuals (rigid concepts) to variables. Only properties can 'escape' syntactic islands. I also give a special interpretation rule for the singular pronoun *one*.

CCT with descriptive pronouns

In the new formalisation that makes us of the set G of *partial assignments* is $\cup\{[W \rightarrow \wp(D)]^X \mid X \subseteq VAR_L\}$. Hence, technically, variables are always assigned properties - but some of these properties represent ordinary objects: the set **D** of *rigid concepts* is defined by:

$\mathbf{D} := \{\mathbf{d} \mid d \in D\}$, where $\mathbf{d} = W \times \{d\}$

Terms are now evaluated as follows:

$$\|x\|^{g,h,w} = \begin{cases} \$(h(x)(w)), \text{ if } x \in \text{dom}(h), \text{ else} \\ \$(g(x)(w)), \text{ if } x \in \text{dom}(g) \text{ and } \$(g(x)(w)) \text{ is defined} \\ \text{undefined otherwise} \end{cases}$$

where $\$(\{u\}) = u$ and $\$(T)$ is undefined, if T is not a singleton set.

I have to change some notational conventions we used earlier in the obvious way:

$S[x:\approx d] := \{\langle g,h',w\rangle \mid \exists h:\langle g,h,w\rangle \in S \ \&\ h[x]h' \ \&\ \$(h'(x)(w)) = d\}$

$S[x] := \{\langle g,h',w\rangle \mid \exists h:\langle g,h,w\rangle \in S \ \&\ h[x]h'\}$

$g[^x/_o] = \{\langle y,o'\rangle \mid y \in \text{dom}(g) \ \&\ g(y) = o' \ \&\ y \neq x\} \cup \{\langle x,o\rangle\}$

$G(S) = \{g \in G \mid \exists h,w: \langle g,h,w\rangle \in S \text{ or } \langle h,g,w\rangle \in S\}$

$GE(S) = \{g \in G \mid \exists h,w: \langle g,h,w\rangle \in S\}$

---

[121] But maybe the example of Kripke is not only special because it makes essential use of a dialogue situation, but also because he uses a definite description as antecedent.

[122] I will do the same later for other constructions that normally are treated as plugs with respect to anaphoric binding.

Also the ordering relation between possibilities has to be changed: <g,h,w> ≤ <g',h',w'> iff w = w' and g ⊆ g' and h ⊆ h'. In terms of this ordering relation, the earlier definition of *substate* can be left unchanged.

For the recursive definition of the context change potential [[A]] [⊆ ℘(G × G × W) × ℘(G × G × W)] of formulae A of **L**, only the interpretation rules for '~A', '∃xA', and 'One(y,x)' are new and have to be given. First, like discussed above, formulae of the form '∃xA' will not only introduce a specific individual, but also a property:

(4) [[∃xA]](S) = {<g'[$^{x}/_{\hat{x}}$|A|$^{g}_{h}$],h',w>|<g,h,w> ∈ S & g ⊆ g' & h ⊆ h' & <g',h',w> ∈ ∪$_{d\in D}$[[A]](S[x:≈ d])},
if ∀g ∈ G(S): x ∉ dom(g), undefined otherwise[123]

The abstraction $\hat{x}|A|^{g}_{h}$ used in the interpretation rule for indefinites is that function f : W –> ℘(D) such that:

$f(w) = \{d \in D \mid [[A]](\{<g,h[^{x}/_{d}],w>\}) \neq \emptyset\}$,

for any w ∈ W.

Second, if we want to account for descriptive pronouns in CCT in a systematic way we have to change also the interpretation rule of negation. What has to be accounted for is that (i) indefinites under the scope of the negation will not introduce specific individuals, (ii) that properties can be introduced by such indefinites, but (iii) that indefinites under the scope of a negation do not introduce properties on the main level that are *dependent* on other terms that stand in monotone decreasing position whose referent is not yet established. For instance, I don't want to introduce properties corresponding with *a woman* in *if a man buys a flower, he gives it to a woman.* because the property introduced by this indefinite in the consequent depends on the referents of *a man* and *a flower* that stand in the antecedent of the conditional.[124] In the interpretation rule below, the first condition is the usual one for negation in CCT and takes care of (i). The second condition takes care of (ii) and (iii). It says that the properties introduced by ~A are those properties introduced by subformulae ∃xB of A that introduce only a single property to the main context:

(2) [[~A]](S) = {<g',h,w>|∃g ⊆ g':<g,h,w> ∈ S & ~∃h" ⊇ h: ∃g" ⊇ g:
<g",h",w> ∈ [[A]](S) & g' = g ∪ {<y,o>| ∃k,h',w':
<k,h',w'> ∈ [[A]]({<l,m,w">|l = g & m = h}) &
y ∈ dom(k)/dom(g) & ∀l,m,n,n':<l,n,w'>, <m,n',w'> ∈
[[A]] ({<l,m,w">|l = g & m = h}): l(y) = m(y) = o}}[125]

---

[123] If indefinites are really ambiguous, referentially used indefinites should be treated as terms that introduce only rigid individuals, and the others as existential quantifiers that introduce only properties. The question how this should be implemented, I leave to the reader .

[124] In section 2.8, however, I will allow these indefinites to introduce functions from worlds and individuals to properties.

[125] Note that it follows that the variable *x* is in the domain of both the first and the second assignment function of a possibility. This is the reason for the somewhat involved interpretation rule for terms.

This definition is rather complicated, but can be simplified if we make use of the following definition:[126]

$\{<x_1,...,x_n>| \psi\}$ in ϕ means $t = \{p \mid \exists y_1,...,y_m \ \& \ p = <x_1,...,x_n> \ \& \ \psi\}$, where $y_1,...,y_m$ are the variables among $x_1,...,x_n$ that are free at t in ϕ.

With the help of this definition we can simplify (2) to (2'):

(2') $[[\sim A]](S) = \{<g',h,w>| \exists g \subseteq g': <g,h,w> \in S \ \& \ \sim\exists h'' \supseteq h: \exists g'' \supseteq g:$
$<g'',h'',w> \in [[A]](S) \ \& \ g' = g \cup \{<y,o>| \exists k,h',w':$
$<k,h',w'> \in [[A]](\{<g,h,w''>|w'' \in W\}) \ \&$
$y \in dom(k)/dom(g) \ \& \ \forall l,m,n,n': <l,n,w'>, <m,n',w'> \in$
$[[A]](\{<g,h,w''>|w'' \in W\}): l(y) = m(y) = o\}\}$

We will keep using the above definition below.

The interpretation rule for 'One(x,y)', finally, is very simple. It works both as an anaphor and a quantifier. It is merely given to prepare the grounds for the analysis of other anaphoric quantifiers later in this chapter.

(6) $[[One(y,x)]](S) = \{<g,h',w>| \exists h: <g,h,w> \in S \ \& \ h[y]h' \ \& \ \$(h'(y)(w)) \in g(x)(w)\}$,
if $\forall g \in GE(S): x \in dom(g)$ and $\forall g \in G(S): y \notin dom(g)$, else undefined

The definitions of truth, entailment, acceptance and acceptability are similar to the ones given earlier.

Let's now discuss a bathroom sentence. Suppose that for each world *w* in W(S), $kard(I_w(P)) \leq 1$, then the following results:

$[[\sim\exists xPx \vee Qx]](S) = \{<g',h,w>|\exists g \subseteq g': <g,h,w> \in S \ \& \ g' = g[^x/_{I(P)}] \ \&$
$\sim\exists h' \supseteq h: \exists g'' \subseteq g: <g'',h',w> \in [[\sim\sim\exists xPx \wedge \sim Qx]](S)\}$

We want to know what $[[\sim\sim\exists xPx]](S)$ is:

$[[\sim\sim\exists xPx]](S) = \{<g',h,w>| \exists g \subseteq g': <g,h,w> \in S \ \& \ I_w(P) \neq \emptyset \ \& \ g' = g[^x/_{I(P)}]\}$

If we assume that in every world in W(S) there was at most one P, the singular pronoun represented as *x* in Qx can be interpreted.

Note that if <g,h,w> is an element of $[[\sim\exists xPx \vee Qx]](S)$, the variable *x* will also be in the domain of g, g(x) will denote in every world all the P's in the world. Is this problematic? Given the definedness conditions on terms and atomic formulae, I don't think so. My theory predicts that you can only use a singular E-type pronoun that is interpreted as *the P*, if in all worlds in W(S) there exists exactly one P. But given that we have used a

[126] Due to Ede Zimmermann (personal communication).

disjunction, and thus that in S it is an issue whether there exist a P, this will typically not be the case.[127]

Note that by the way E-type pronouns are interpreted, also clauses of the form "$\sim\exists xP(x,y) \vee Qx$" can be interpreted in context S, if for all $\langle g,h,w\rangle \in S$ it is the case that $y \in dom(g) \cup dom(h)$ and there is at most one $d \in \mathbf{D}_w$ such that $\langle d, \|y\|^{g,h,w}\rangle \in I_w(P)$.

This is in general the case. And needed too, to account for the fact that E-type pronouns can be *relational*, and *indexical* (cf. Neale, 1990):

(17) *Smith's murderer* is insane. *He* should be jailed for life.
(18) The one who will win *this* game will be lucky. *He* will get all the money.[128]

I formulated the approach for indefinites. But of course, you can also refer back with a singular pronoun to a *definite description* where original CCT predicts this to be impossible. Just like indefinites, they introduce a concept. Special about non-anaphorically used (singular) definite descriptions is that in every world of the context resulting after the interpretation of this description there will be only one object that satisfies the description. This has two consequences. First, the concept that is introduced can be restricted to the noun phrase itself. Second, it is assured that you can always refer back to such a description if you are considering a world in the same context as in which the description was used. I propose that a sentence like *The N is P*, where *the N* is used non-anaphorically, is represented as follows: $\iota x[Nx] \wedge Px$, and that in general ιxA is short for '$\exists x \forall y[A[^x/_y] \leftrightarrow y = x] \wedge A$', where $A[^x/_y]$ is A with all occurrences of free x replaced by fresh y.

Thus, the theory predicts that singular pronouns can sometimes pick up definite descriptions introduced in positions predicted to be inaccessible by standard CCT. Here are some examples that suggest that this is needed:

(19a) If John makes coffee, *his wife* will be happy. *She* is a nice person. (after V.d. Sandt, 1992)
(19b) If all countries have presidents, *the president of France* probably regards himself as their cultural leader. *He* is such a pompous ass. (Geurts, 1995)

These examples ware supposed to show that presuppositions triggered in the consequent of a conditional do not give rise to a conditional presupposition. How else could we interpret the unbound pronoun in the second sentence? The argument is a forceful one if based on the assumption that negations are absolute plugs with respect to anaphoric binding. But on the motivation for CCT that I have given, the argument loses its force. Not the conditional gives rise to the presupposition that John has a wife, or that France has a king, but the sentence in which the E-type pronoun occurs. Note that if in the antecedent of (19a) we substituted *a man* for *John*, the description *his wife* could not be anaphorically picked up by a subsequent sentence, which is indeed what I predict.

By our use of two assignment functions, we can also account for the fact that proper names used in positions that make introduced variables not accessible according to standard CCT

---

[127] But sometimes it is, as in *Either there is no bathroom in the house, or it is in a funny place. In any case, it is not on the ground floor.* Note that for this example it is crucial that the pronoun in the second sentence is a descriptive pronoun, and that the description must have smaller scope than the negation.

[128] Neale argues that a lot of incomplete definite descriptions can be completed by purely referential or indexical material. I agree that this is natural to assume for descriptions like *the mayor* or *the murderer*, but I don't think that descriptions should normally be treated as Russellian descriptions (see also Evans, 1982, pp. 324-325)

can always be picked up in the ungoing conversation. We don't need a special proper name rule to account for this if we assume that proper names should be treated as rigid designators.

Note that by our interpretation rules it is predicted to be possible that sometimes a pronoun can pick up a description that is interpreted in a world, or a more complex index, that is not an element of the set of indices of the context resulting after the interpretation of the indefinite. This is good news. Examples are easy to find where non-rigid concepts seem to be useful:

(20) *My home* once was in Maryland, but now *it*'s in Los Angeles.(Partee, 1972)[129]
(21) Senator Green believed that he had nominated *the winner of the election*, but Senator White believed that she had nominated *him*. (Partee, 1972)[130]
(22) This year *the president* is a Republican.
Next year *he* will be a Democrat. (Evans, 1977)
(23) John believes that *the winner of the game* needs to play well, while Mary believes that *he* just must be lucky.

Can we treat all anaphoric dependencies across attitude verbs even with more agents involved by this implementation of the E-type approach in CCT? That will be an issue for the next chapter.

2.7 Epistemic *might* and modal subordination

I argued above that descriptive pronouns are useful to account for anaphoric relations that standard CCT cannot account for in cases where no explicit modal is mentioned. In this section I will argue that descriptive pronouns are also useful in case such explicit modals are mentioned, but that we need something extra, too. In this section, I will only consider the modal *might*. Both Kibble (1994) and Geurts (1995) have recently argued that there are quasi-anaphoric relations between uses of modal verbs in discourse. The argument is that the second sentence of a discourse like (24) is ambiguous:

(24) It might be that a thief broke into the house.
And it might be that the alarm was going of.

The embedded clause of the second sentence can be interpreted under the assumption that a thief broke into the house, and it can be interpreted under another assumption. In the first case, the second use of *might* is 'anaphoric' to the first use of *might*, and in the second case it is not. To account for this quasi-anaphoric dependence, we can add propositional discourse markers to our variables, and represent a sentence like *It might be that A* by '$\Diamond^{p}_{q}$A' where variable *p* represents the information with respect to which the *might* operator is interpreted, and variable *q* represents the information represented under *p* updated by A. If *g* is an assignment function that has *p* in its domain, the question is what kind of information state g(p) should denote. Geurts (1995) assumes that to account for the anaphoric dependencies in (25)

(25) *A thief* might break in. *He* might steal the silver,

---

[129] Such an example shows not only that non-rigid concepts are useful, but also that analysing pronouns as descriptive pronouns can only be done appropriately in a dynamic framework. The pronoun *it* in (20) has a different meaning than the pronoun used in (20):
(20') *My favourite restaurant* once was in Maryland, but now *it*'s in Los Angeles.

[130] Partee (1972) notes that (21) is ambiguous. Either the two senators dispute over who nominated a certain person, or over who the winner of the next election will be, the one nominated by Green or the one nominated by White. She concludes that therefore a sentence like (21) "constitutes a real problem for any attempt to find a uniform basis for the pronoun-antecedent relationship" (p. 425).

we have to assume that what is denoted by g(p) is a set of world-assignment pairs. I will follow Kibble (1994), however, and assume that g(p) just denotes a set of possible worlds. So, just like indefinites introduce (also) a property, epistemic *might* introduces a proposition. My argument will be that pronouns that stand under the scope of a *might* operator and that anaphorically depends on an antecedent that also stands under the scope of a epistemic modal operator should be treated as descriptive pronouns. The argument will closely resemble the argument against the approach of Krahmer & Muskens (1995) given above; also here the pronoun should be treated as a descriptive pronoun because it is uniqueness that counts.[131] Consider the following minimal pair due to Ede Zimmermann (personal communication):

(26) It is likely that Hans has *an eyedoctor* in London.
And it is possible that he has seen *him* recently.

(27) It is likely that Hans knows *a doctor* in London.
*?And it is possible that he has seen *him* recently.

Intuitively, I think, the use of the singular pronoun *him* in the first discourse is o.k., but this is not the case for the second discourse. I want to suggest that this difference in acceptability is due to our world-knowledge that people have usually at most one eyedoctor, but that everybody usually knows more than one doctor.

We can model the information states that epistemic *might* anaphorically refers back to by sets of possible worlds, if in such cases descriptive pronouns are always relevant.[132] In general, sentences like *It might be that A* , and *It is possible that A* will be represented by something like '$\Diamond^{p}_{q}$ A' and interpreted as follows:

$$[[\Diamond^{p}_{q} A]](S) = \{\langle g',h',w\rangle \mid \exists g,h:\langle g,h,w\rangle \in S \,\&\, g \subseteq g' \,\&\, \exists k,w': \langle g',k,w'\rangle \in [[A]](\{\langle g,m,w''\rangle \in S \mid w'' \in h(p)\}) \,\&\, h[q]h' \,\&\, h'(q) = W([[A]](\{\langle g,m,w''\rangle \in S \mid w'' \in h(p)\}))\}$$

where $W(S) = \{w \in W \mid \exists g,h: \langle g,h,w\rangle \in S\}$

In this way we cannot only account for the difference in acceptability between (26) and (27). It can also account for the acceptability of (15):

(15) It is possible that John does not own *a donkey*,
but it is also possible that he keeps *it* very quiet.

and it can be explained why the following discourse is unacceptable:

(28) Andy might have *a bike*. **It*'s a Harley.

Sentence (15) is acceptable because *it* simply stands for *the donkey that John owns*. The use of a singular pronoun in the second sentence of (28) is odd because, according to my explanation, the pronoun must be a descriptive pronoun, and singular descriptive pronouns can only be appropriately used if in the worlds of all possibilities of the context the description has a unique instantiation. But this latter condition will typically not be satisfied

---

[131] Kibble (1994) does not motivate his account in this way.

[132] I would say the same for modal subordination cases with examples of conditionals. The contrast between the following two discourses discussed in Roberts (1989) can be accounted for by descriptive pronouns: (a) *If I had the money, I'd hire a gardener. I'd pay him $10 an hour. (b) ?If I had the money, I'd hire a gardener. I pay him $10 an hour* (cf. Neale, 1990, p. 189). See also chapter 6 for a similar analysis of attitude attributions conditionally dependent on desire attributions.

in a conversational situation in which the first sentence of (28) is uttered, because in those situations it is normally also consistent with what is presupposed that Andy does not have a bike.

Note that in the above interpretation rule for epistemic *might* I made implicitly three claims. First, that epistemic *might* introduces a proposition represented by a set of possible worlds that can be taken up anaphorically by later occurrences of such a modal. Second, that indefinites inside the scope of epistemic *might* are not referentially used and only introduce properties. Third, that epistemic *might* quantifies not only over worlds, but also over assignments. The first two claims were directed against Geurts' analysis of modal subordination.[133] The third claim is in accordance with the analysis of epistemic *might* by Groenendijk *et al.* (1994) that I discussed in a foregoing section. Still, it is interesting to see to what extend we can give an alternative analysis of epistemic *might* once we have descriptive pronouns around. The alternative rule will look as follows:

$$[[\Diamond^{p}_{q} A]](S) = \{<g',h',w> | \exists g,h:<g,h,w> \in S \;\&\; g \subseteq g' \;\&\; \exists k,w': <g',k,w'> \in [[A]](\{<g,h,w> \in S | w \in h(p)\}) \;\&\; h[q]h' \;\&\; h'(q) = W([[A]](\{<g,h,w> \in S | w \in h(p)\}))\}$$ [134]

This interpretation rule is different from the one given earlier in that epistemic might is no longer treated as an unselective quantifier. In the above interpretation rule it quantifies only over worlds. That is, in the above interpretation rule for epistemic *might*, I evaluate the modal distributive with respect to the assignments, and only globally with respect to the worlds.[135] As a result, epistemic *might* only expresses uncertainty about the subject matter of conversation.

In terms of Dekker's subjects, we might say that the solution of Groenendijk *et al.* comes down to this: if the speaker says *Someone committed the murder. It might be the butler*, the speaker is saying that he is not clear about the identity of the subject introduced by the first sentence. The subject that corresponds with it is not a rigid function (with respect to the context). Note that a subject as Dekker defined it, looks very much like an individual concept. It's a function from possibilities to objects. But the crucial difference is that it's a more fine grained object, because possibilities in a CCT information state are more fine grained than possible worlds.

If the second interpretation rule for epistemic *might* is correct, we don't need this extra fine-grainedness. There is a strong intuition behind the second interpretation rule: if an indefinite is referentially used by a speaker, he will normally not communicate his uncertainty of the identity of this referent by means of epistemic *might*. And if he does, he will express his uncertainty whether this individual satisfies a certain uniquely identifying description.[136]

Whether this intuition is right or not, at least we have to see whether we can account for Dekker's problem and the donkey-closet problem if the second interpretation rule is assumed. First, note that on the above interpretation rule for *might*, Dekker's problem does not arise. Contrary to unbound pronouns, bound pronouns are never interpreted as descriptive pronouns. Because we have assumed with Groenendijk *et al.* that a sentence

---

[133] See chapter 3 of this dissertation for a more general discussion of Geurts' approach.

[134] Note that according to this interpretation rule pronouns inside the scope of the *might* operator need not be descriptive pronouns in case it has an indefinite as antecedent that is 'introduced' in the 'main' context.

[135] In this I follow Van Eijck & Ceparello (1994). They don't make use of descriptive pronouns, however, and assume that variables are rigid expressions. As a result, they cannot account, for instance, for the distinction between '$\Diamond x = y$' and '$x = y$'.

[136] According to Fodor (1979), in all appropriately used identity statements of the form t = t', one of the two terms must be used referentially and the other attributively. Remember also the Stalnaker (1981) solution of self locating beliefs like *I am Lingens*.

represented by $\exists xA$ should not be interpreted as the conjunction of $\exists x$ and A, Dekker's problem is solved in the same way as they solved it.

What about the donkey-closet problem? Let's consider the examples and the model again that motivated the proposal of Groenendijk *et al.* Note that there is something special about this model, in all the relevant worlds there is only one person hiding in the closet. On our account, in these cases the pronoun might be used as a referring expression, but need not. It can also be used descriptively, as an E-type pronoun. As E-type pronoun it refers to *the person who is hiding in the closet.* In this way the non-equivalences of (7) versus (8), and (9) versus (10)

(7) There is someone hiding in the closet. He might be the one who did it.
(8) There is someone hiding in the closet who might be the one who did it.
(9) If there is someone hiding in the closet, he might be the one who did it.
(10) Anyone who is hiding in the closet might be the one who did it.

are explained by assuming that the unbound pronouns in (7) and (9) can be interpreted as E-type pronouns. Also in this way it can be explained why (7) and (9) are less informative than (8) and (10), respectively. Note that by allowing unbound pronouns to be interpreted as E-type pronouns, we predicted already, just like Groenendijk *et al.* do, that the donkey equivalences $\exists xA \wedge B \Leftrightarrow \exists x[A \wedge B]$ and $\exists xA \rightarrow B \Leftrightarrow \forall x[A \rightarrow B]$ are not valid anymore.

This seems like a nice picture. Still, Groenendijk *et al.* (1994) have given an example where the extra fine-grainedness of subjects seems to be needed anyway. Consider the following mathematical example:

(29) $\exists x[x^2 = 4] \wedge \Diamond x = 2 \wedge \Diamond x = -2$

They claim that this is an example that illustrates why we need to quantify over assignments, because the first claim doesn't give us new information about the world, in every world 4 is a square of 2 and -2. But I'm not completely convinced by this example. I find the discourse "There is a number whose square is 4. It might be 2, and it might be -2" very unnatural. I find it much more natural that a speaker would say something like "There is a number on this card whose square is 4. The number might be 2, and it might be -2. Can you guess what number it is?" But in this case the definites *the number* and *it* can simply stand for the description *the number on this card* and would be no counterexample to the second way of interpreting epistemic *might.*

2.8 Functional pronouns and arbitrary objects

Until now we have accounted for (co-) referential and descriptive singular anaphora. But singular pronouns can also be used when they do not co-refer with the speaker's referent of an indefinite used in a foregoing clause, or as an abbreviation of a definite description. Sometimes they function as an abbreviation of a *possessive.* Note that with the machinery introduced in § 2.6, I cannot yet account for Karttunen's (1969) paycheque example: *The man who gave his paycheque to his wife was wiser than the man who gave it to his mistress*, although it is usually assumed that the E-type approach is appropriate here, too.[137] The reason I cannot yet account for these examples is obvious; the functions I introduce are functions from worlds to sets of individuals, in general, however, I should introduce functions from world-sequence pairs to sets of individuals. I conjecture that this can be easily done, but I will only show it for the case where the sequence of individuals consists of at most one individual. In § 2.6 we have accounted for the case where the sequence is the empty sequence; then the pronoun goes proxy for a definite description

[137] see Chierchia (1996), for instance.

recoverable from the antecedent clause. In this section I want to account for the case where the sequence consists of one individual; the case where the pronoun goes proxy for a possessive recoverable from the antecedent clause. In order to do this, I use the abstraction operator also used in § 1.14. Just as in chapter 1, I will say that if A is a formula and *x* a variable, $\hat{x}A$ is a one-place predicate. The idea is that possessives like *the paycheque of* are represented by something like $\hat{x}(\iota y[\text{Paycheque\_of}(y,x)])$, and that pronouns can be represented by something like f(x), where *f* denotes a function from individuals to sets of individuals in each world, and *x* denotes an individual. We need to know two things: (i) how should complex predicates like $\hat{x}A$ and atomic formulae like $\hat{x}A(y)$ be interpreted? and (ii) how will our more complex functions be introduced?

The problem with the former question is that a complex predicate like $\hat{x}A$ may be build up from a formula A that introduces new individuals. Thus, the interpretation rules of complex predicates and of atomic propositions containing complex predicates have to be *internally dynamic*. This is assured by the following two interpretation rules, of respectively complex predicates, and of atomic formulae built up by a complex predicate:

$$[[\hat{x}A]](S) = \lambda d.[[A]](S[x:=d])$$

$$[[\hat{x}A(y)]](S) = \{<g',h',w> | \exists g \subseteq g', \exists h \subseteq h': <g,h,w> \in S \;\&\; <g',h',w> \in ([[\hat{x}A]](S)(||y||^{g,h,w}))\}$$ [138]

Thus, complex predicate $\hat{x}A$ denotes the function from a context *S* and an individual *d* to the interpretation of A in S, if *x* is assigned to *d*. The atomic formula $\hat{x}A(y)$ is then basically the application of this function to the interpretation of *y*, for each possibility of the context.

The answer to the second question is now straightforward. Formulae of the form '∃xA' do not just introduce properties, but functions from worlds and n-tuples of objects to sets of individuals. As mentioned before, I will limit myself to the case where this n-tuple is at most one object. What is introduced will no longer be represented by '$\hat{x}|A|^{g}_{h}$', but will be represented by '$\hat{x}|A|^{g}_{h},s$', where *s* is a possibly empty sequence of variables occurring as free variables in A that are bound by an abstractor.[139] Let us consider the case where *s* is the sequence <y>, the case where variable *y* is the only free variable in A that is bound by an abstractor, like in '$\hat{y}(\exists xR(x,y))$'. The denotation of '$\hat{x}|A|^{g}_{h},<y>$' will now be the function *f* $W \times D \rightarrow \wp(D)$ such that

$$f(<w,d>) = \{d' \in D | [[A]](\{<g,h[^{x}/_{\mathbf{d}'}, {}^{y}/_{\mathbf{d}}],w>\}) \neq \emptyset\}$$

for every $w \in W$, where $\boldsymbol{d}$ denotes the rigid individual concept associated with *d*.

So, the intensional function introduced under variable *x* by $\hat{y}(\iota x\text{Paycheque\_of}(x,y))$ corresponds with *the paycheque of*, and if a singular pronoun takes this possessive as antecedent, like the pronoun in *the man who gave it to his mistress*, it refers to the unique paycheque of the man, if applied to this man.[140] This solution of paycheque examples

---

[138] If it is assumed that also terms can have dynamic effects, the interpretation rule has to be changed slightly. But this is straightforward, and left to the reader.

[139] This means that we have to keep track of the variables that are bound by an abstractor. I will assume that this can be done.

[140] That is, if the pronoun *it* has a *sloppy* reading. To account for the *strict* reading, the function associated with *the paycheque of* is applied to the original man. But it is better to introduce both a normal individual, and a function from individuals to properties. This can be accounted for by representing a possessive like

presupposes that singular pronouns should not always be represented by individual variables, sometimes they will be of the form 'x(y)', meaning that the pronoun refers to the denotation of *x* applied to the denotation of *y*.

Functional pronouns cannot only be used for the analysis of 'paycheque' examples, but will also be helpful to account for the following two cases:

(30) Every man lost a pen, and some man found *it*.
(31) Every player chooses a pawn. *He* puts *it* on square one. (Roberts, 1989)

Intuitively, the pronoun *it* in (30) refers to the unique pen that the man referred to by *some man* in the second conjunct lost. In (31), *he* refers to any arbitrary player who chooses a pawn, and the pronoun *it* to the unique pawn that this player has chosen. As it stands, our CCT can neither account for (30) nor for (31). In fact, proponents of *Constructive Type Theory* have argued that their framework is to be preferred above something like our CCT, because it can account for sentences like (31), and the philosopher Kit Fine, who formalised the theory of *arbitrary objects*, argued for his approach, because it can account in a very simple way for sequences like (31). But of course, the fact that our current version of CCT cannot handle (30) and (31) doesn't mean that CCT cannot be extended in such a way that it can handle both kinds of anaphoric relations.[141]

To account for the anaphoric dependencies in (30) and (31) we have to assume that also sentences of the form *Every A, B* will introduce some kinds of objects. I will argue that these sentences will introduce two kinds of objects needed to refer back with *singular pronouns* to expressions in such a sentence. First, the sentence will introduce *an arbitrary object* that (arbitrarily) refers to any A that is B.[142] This arbitrary object will simply be the representative of the property *A that is B*. If a singular pronoun refers back to this introduced object in another sentence, this other sentence can only be true if the sentence is true for any A that is B. I will use the theory of arbitrary objects in only a very limited way, I won't make use of the *dependence relation* that is so crucial in Fine (1983). But how then will the pronoun *it* be interpreted in (30) and (31)? Here I will propose that also indefinites used in the scope of a universal quantifier can introduce functions from world-sequence pairs to sets of individuals. For instance, the indefinite *a pen* in *Every man lost a pen* introduces a function in intension, f, corresponding to *pen of*. The second conjunct of (30) will then be represented by something like $\exists x[\text{Man}(x) \wedge \text{Found}(x, f(x))]$.[143] Similarly, if x is a variable that refers to an arbitrary player, the second sentence of (31) will be represented by something like P_S1(x, f(x)). Because *it* is a singular pronoun, f(x) will be only interpretable when it is assumed (maybe after accommodation) that every man lost only one pen in (30), and that every player chooses only one pawn in (31).[144] To be able to introduce such a function, I will represent a universally quantified sentence by something

---

Bill's phone number by $\bar{y}\iota z_v$PhoneNo.-of(z,y)(b). Bill's phone number is now introduced under z, and the function under v.

[141] Roberts (1989), Groenendijk & Stokhof (1990) and Dekker (1993) have accounted for sentences like (31) by claiming that sometimes discourses that are normally formalised by $\forall x(A,B) \wedge C$ should sometimes be interpreted as $\forall x(A, B \wedge C)$. I resist this proposal for the usual reason; in this way we cannot account for the uniqueness implication of singular anaphora.

[142] Such an arbitrary object will only be introduced by 'singular' universal noun phrases like *every*, not by every's 'plural' counterpart *all*.

[143] I will assume that such a translation can somehow be given in a systematic way.

[144] This is the reason why I think such a solution is to be preferred above accommodation accounts (Roberts (1989)), or accounts where a dynamic negation is introduced (as in Groenendijk & Stokhof (1990) and Dekker (1993)). For a radically different approach towards sentences like (30) and (31), see Fernando (1994). Unfortunately, also Fernando's approach cannot explain the uniqueness condition I seek to implement.

like $\forall x(A, B)$. Thus, the first conjunct of our sentence (30) is then represented as $\forall x($ Man(x), $\hat{y}(\exists z[\text{Pen}(z) \wedge \text{Lost}(y,z)])(x))$.

With respect to the *syntax*, first we extend the language with variables for function symbols, and then we have to make a distinction between three kinds of terms, individual variable, variables standing for arbitrary objects, and terms like f(t), where *f* is a variable denoting a function, and *t* is any kind of term. Thus, for simplicity I will assume that we have to extend our language with different kinds of variables. Some variables stand for properties, some for functions from worlds to arbitrary objects (the set AVAR), and others for functions from worlds and individuals to sets of individuals. As a consequence, assignment functions can no longer be thought of as functions from variables to properties, but have to be defined as functions from variables to either properties, or arbitrary objects, or functions. *Models* are five tuples <W,D, {**A**(w)| w ∈ W}, {V^w| w ∈ W}, I>, where W, D and I are as before, for every world *w* **A**(w) is the set of arbitrary objects in *w* disjoint from D(w), and for every *w*, V^w is a set of *value assignment functions*, a set of functions from **A**(w) to D(w). I will assume that if *A* is a sentence, *x* a variable, and *g* and *h* assignment functions, there is a function **a** from worlds to arbitrary objects such that for all *w* in W, {v(**a**(w))| v ∈ V^w} = $\hat{x}|A|^g_h$(w). I will abbreviate this function by a-$\hat{x}|A|^g_h$.

Having extended the language with new terms, we now have to give a new interpretation rule for these terms:

$[[t]]^{g,h,w,v}$ = $(g(t)(w))$, if t ∈ dom(g) & $(g(t)(w))$ defined, else
= $(h(t)(w))$, if t ∈ dom(h) & $(h(t)(w))$ defined,
= v(g(x)(w)), if t ∈ AVAR & t ∈ dom(g)
= $(g(f)(w)([[t']]^{g,h,w,v}))$, if t = f(t') & $(g(f)(w)([[t']]^{g,h,w,v}))$ defined,
= undefined otherwise

It is a well known consequence of the theory of arbitrary objects that the wanted semantic value of a whole sentence cannot functionally depend in the usual way on the corresponding semantic values of its parts. In our case the problem is that the wanted semantic value of the second sentence of a problematic discourse like *Every natural number is greater or equal to 0. It is even or odd* should be that every natural number is even or odd. But if the semantic value of the disjunction is determined from the *same kind* of semantic values of its parts, the semantic value of the whole sentence would be saying that every natural number is even, or every natural number is odd. Fortunately, this problem has a familiar solution. Just like in chapter 1 we defined the n-ary notion of truth of a sentence in terms of the n+1-ary notion of truth with respect to a counterpart function by means of supervaluation, now we define the wanted semantic value of a sentence in terms of the semantic value of a sentence with respect to a valuation function. With respect to the interpretation rules, I will now introduce a new interpretation function, [.], that is defined for all natural language sentences, but not for all clauses, in terms of [[.]]. For any formula *A* that represents a natural language sentence, [A] is defined as follows:

$$[A](S) = \{<g',h',w>|\ \exists g \subseteq g', h \subseteq h' : <g,h,w> \in S\ \&\ \forall v \in V^w : <g',h',w,v> \in [[A]](\{<k,l,w',v'>|\ <k,l,w'> \in S\ \&\ v' = v\})\}.$$

The interpretation function [[ ]] will also be changed, but except for the clause of universal quantification, in the almost trivial way. Atomic formulae will be interpreted now as follows:

$$[[Pt_1 \ldots t_n]](S) = \{<g,h,w,v> \in S \mid <|\!|t_1|\!|^{g,h,w,v}, \ldots, |\!|t_n|\!|^{g,h,w,v}> \in I_W(P)\},$$
$$\text{if } \forall x_i\colon 1 \leq i \leq n\colon \forall <g,h,w,v> \in S\colon |\!|t_i|\!|^{g,h,w,v} \text{ is defined, undefined otherwise}$$

The new interpretation rule for universal quantification will now be more complicated. As motivated above, a sentence of the form *Every A has a B* is not only statically checked for truth, but also has a dynamic effect because two new objects are introduced, namely a function from worlds to an arbitrary object, *the arbitrary A that has a B*, and the function in intension from A's to B's owned by the A.

$$[[\forall x(A, B)]](S) = \{<g'[^{x}/_{a\text{-}\hat{x}}] \mid A \wedge B|^{g}_{h}, h, w, v> \mid <g,h,v,w> \in S \;\&$$
$$\{d \in D \mid [[\hat{x}A]](\{<g,h,w,v>\})(d) \neq \emptyset\} \subseteq$$
$$\{d \in D \mid \exists h'\colon <g', h', w, v> \in [[\hat{x}(A \wedge B)]]\,(\{<g,h,w,v>\})(d)\}\}$$

I think it is important to see how arbitrary objects and functions corresponding to possessives can be introduced. Still, for reasons of convenience, I will mostly ignore functional pronouns and arbitrary objects in the rest of this dissertation. I just wanted to point out that it is possible to account for paycheque examples, and for the anaphoric dependencies in (30) and (31), in a way that uses certain formal tools that are very similar to the tools that can already be used to account for E-type pronouns. Still, the use of functional pronouns will be used in §3.9 to account for Edelberg's asymmetry problem, and will be essential again in § 4.11 to account for presuppositions in quantified contexts in a two-dimensional framework.

2.9 Anaphoric quantifiers, quantified pronouns and salience[145]

A speaker can use a singular pronoun appropriately because in every possibility of the context there is an object available for reference to which this pronoun refers. There are two ways why an object is salient, or available for reference. Either it is salient because it is observable for both speaker and hearer, or because (normally) the speaker made it salient by using an indefinite description. Obviously, not only single objects are available for reference in this way, a *set* of objects can become salient in one of those ways, too. There might for instance be obvious criteria to select subsets of observable entities in the environment that speaker and hearer share, and the speaker can make a set of individuals salient by using a plural indefinite. We can refer to such sets by plural pronouns. For the sets that are available for reference by observable criteria we have deictic and demonstrative uses of the plural pronoun *they*, and for the anaphoric, but non-E type, use of *they* we have already seen the example of Dekker (1994) that I repeat here:

(4) Yesterday, John met some girls. They invited him to their place

Plural pronouns can refer back to salient sets, but these sets can have another function, too. If a speaker says *Everybody had a good time*, he probably is not claiming that everybody in the universe had a good time. He restricted his domain of quantification to a certain set of individuals. For the assertion to be understandable for the hearer, this set of individuals must be salient somehow. It's natural to assume that such a quantifier can restrict its domain of quantification by the same sets that are available for reference for plural pronouns. This suggests that the interpretation of quantified noun phrases in a possibility of the context depends in the same way on the reference-context of that possibility as plural pronouns. The domain of quantification can be determined either by deictic or by anaphoric means. Quantifiers are not two-, but three-place relations. Indeed, this has been proposed

[145] This section is crucial for the analysis of presuppositions in quantified contexts that will be proposed in chapter 4.

by Westerstahl (1984) and Van Deemter (1991).[146] Because in certain situations a sentence like *All cheered* makes sense, a salient context set is sometimes needed to be able to interpret the sentence in which a certain anaphoric quantifier (determiner) occurs. Westerstahl even showed that we need the existence of several salient contexts sets apart from the domain of discourse to give a reasonable interpretation of sentences in which more quantifiers occur.[147] Consider the following example of Westerstahl:

(32) The English love to write letters. Most children have several pen pals in many countries.

To make sense of this sentence, we have to assume that the domain of discourse must contain both English and non-English children, although *most* is restricted to the set of English children, but *several* is not.

Let us, for simplicity, call a quantifier whose interpretation depends on a salient context set an *anaphoric quantifier*. Of course, if we assume that the domain of discourse is always a salient context set, we can claim that all quantifiers are anaphoric. But how much content really has this claim? Can indefinites ever be anaphoric to a context set not equal to the domain of discourse, and what does it mean to claim that proper names, treated as generalised quantifiers, are anaphoric? With respect to the first question, consider the following example due to Van Deemter (1991):

(33) A herd of elephants was visible in the rear window. *Two elephants* were lying somewhere in the middle.

Obviously, the speaker would use the indefinite *two elephants* to refer to two elephants visible in the rear window. By the use of the indefinite he selects a subset of the already salient group of elephants, and make this new subset salient, too. That is, we don't have to use an E-type pronoun to refer back to this new introduced set. I believe that this is in general the case for quantifiers that are leftward monotone increasing. Where the use of *a man* makes an individual available for reference in each possibility of the context, the use of *some men* (*two men*) makes a set of (two) men salient in each of the possibilities.

In the way I motivated CCT, there is a distinction between indefinites and quantifiers. Indefinites are referential expressions that make new objects available for reference, while quantifiers don't. Quantifiers only introduce properties to which we can refer back to by E-type pronouns. What should we think of incomplete definite descriptions?

A speaker can use a singular pronoun appropriately, because he intends to refer to an object that has somehow become available for reference. How else would it be able to refer to a single object? Strawson (1950) observed that something similar is true for so called *incomplete definite descriptions* like *the table*. There is more than one table in the universe, so the noun phrase cannot determine a unique object by the descriptive material itself. But then, how could a speaker appropriately use this noun phrase to refer to a unique table? Strawson (1950) suggested that this could be true because the incomplete description is referentially used, and some proponents of original CCT have assumed the same. An incomplete definite description must refer to a salient object. In original CCT it was assumed that the reference of incomplete descriptions should be determined by co-indexing. On this picture it follows that on the referential reading the descriptive material is at most relevant for reasons of appropriateness. However, this is not what we find. Also anaphorically used incomplete definite descriptions can select an object from a bigger set, thereby making a new object available for reference:

---

[146] Van Deemter argues for a much more general conception of anaphora. I only try to account for what he calls *subsectional anaphora*, but we have other kinds of anaphora, too. I can't do full justice to his more general view.

[147] It can be assumed that the domain of discourse is always salient.

(34) Two men are walking in the park. The tallest man is wearing a hat.
(Groenendijk *et. al.*, 1995b)

The example shows that information about the subject matter of discourse should not be represented separately from the information about the discourse itself. With respect to proper names, already Sommers (1982) claimed that they should be treated as anaphoric quantifiers. And in fact proper names can both select an object from a bigger set, and at the same time make salient an object available for reference by a pronoun or short description. Consider the following updated variant of an example of Weijters (1989):

(35) Last year, Ajax did not succeed in winning the cup final against Juventus, although Littmanen scored a beautiful goal.

Even a hearer who is not interested in European soccer would be able to infer that Littmanen, whoever that might be, is a member of the Ajax team.

If proper names can select an object from a bigger set, it is to be expected that even pronouns can do the same. And, indeed, this is what we find:

(36) Our neighbours are extremely nice people.
He is a teacher, she is a housewife. (Hendriks & Dekker, 1996)[148]

We have seen that it is not only a good idea to treat quantifiers as special kinds of anaphora, but also that there are good reasons to treat definite descriptions and pronouns as special kinds of quantifiers. The reason to treat descriptions and pronouns as quantifiers is not just that they can make new objects available for reference. That only shows that they have something in common with indefinites. By treating them as quantifiers and not as terms, it also gives us the possibility to let them stand in non-trivial scope relations with logical operators. That this is needed is shown by the E-type pronoun *they* in the following discourse:

(37) Most friends of Sue will marry a Swede. Sue believes *they* will be happy

The second sentence has a reading where the description *the friends of Sue who will marry a Swede* abbreviated by *they* is interpreted in a *de re* way, but also one where the description is interpreted *de dicto*.[149] Similarly, the descriptive pronoun *he* in the following example

(38) *The mayor* is a democrat. John thinks that next year *he*'ll be a republican.
(Neale, 1990, p. 214)

can take intermediate scope. The most familiar way to account for those facts is to treat pronouns as quantifiers.[150]

It is only natural to expect that the scope of a quantified pronoun matters if the pronoun is interpreted descriptively, if it is an E-type pronoun. But scope could even be relevant for

---

[148] Hendriks & Dekker hypothesise that an anaphor can only select a part of a bigger set, behave as a non-monotone anaphor as they call it, if it has in English a special kind of accent that marks topic- or linkhood.

[149] Evans (1977) cannot account for the *de dicto* reading, because he treats E-type pronouns as referring expressions. Looking only at an extensional fragment, Kamp & Reyle (1993) assume that quantifiers introduce sets, and not properties. Our example shows that when the fragment is extended with verbs like *believe*, this won't be good enough.

[150] But see chapter I for an alternative. There I treated definite descriptions as singular terms whose referent is world dependent. I still could account for scope differences by means of the abstraction operator. To account for plural pronouns we can then use an optional distribution operator.

quantified pronouns that anaphorically refer back to a *rigid* entity. The reason is that this rigid entity can correspond to a non-singleton set. A sentence like *They walk* is true just in case everybody in the relevant context set walks. Similarly, in the most natural reading of *They don't walk* the sentence seems to be true just in case none of the individuals in the picked up context set walks. How can we account for this reading? Dekker (1994) suggested that to account for plurals in CCT, we have to resort to truth-conditionally partial semantics. Happily enough we don't have to, we don't need partial semantics for this purpose. If we treat pronouns as quantifiers, we can simply say that the pronoun *they* has wide scope with respect to the negation. Thus, the scope of a quantified pronoun even matters in extensional contexts.

If we want to determine the meaning (ccp) of a sentence in a compositional way from the meanings of its parts, quantifiers will denote functions from properties to ccp's. Quantifiers come with two variables, the variable that denotes the set on which the quantifier is anaphorically dependent, and the variable which it binds and introduces. The quantifiers *a man, one, most men, every man, they, the man, he*, and *John* can then be represented as follows (where x is a numberless pronoun, properties can be introduced under all kinds of pronouns[151]):

some$^{x}_{y}$ men	=	$\lambda Q{:}\exists^{x}_{y}(\mathrm{Man}(y),Qy)$
a$^{x}_{y}$ man	=	$\lambda Q{:}\exists^{x}_{y}(\mathrm{Man}(y),Qy)$
one$^{x}_{y}$	=	$\lambda Q\exists^{x}_{y}(\mathrm{T}(y), Qy)$, where T is the trivial property
most$^{x}_{y}$ men	=	$\lambda Q{:}\mathrm{Most}^{x}_{y}(\mathrm{Man}(y),Qy)$
every$^{x}_{y}$ man	=	$\lambda Q{:}\forall^{x}_{y}(\mathrm{Man}(y),Qy)$
they$^{x}_{y}$	=	$\lambda Q{:}\forall^{x}_{y}(\#(x) > 1 \wedge \mathrm{Ty},Qy)$,[152] where $\#(x) > 1$ means that the cardinality of the denotation of x is greater than 1.[153]
the$^{x}_{y}$ man	=	$\lambda Q{:}\iota^{x}_{y}[\mathrm{Man}(y)] \wedge Qy$
he$^{x}_{y}$	=	$\lambda Q\iota^{x}_{y}[\mathrm{Male}(y)] \wedge Qy$
John$^{x}_{y}$	=	$\lambda Q{:}\iota^{x}_{y}[y = \mathrm{John}] \wedge Qy$[154]

I think it is useful to introduce two special kinds of variables to which quantifiers can anaphorically refer back. First, I introduce the designated variable α. This variable always refers back to the universe of the respective worlds. Second we have variable δ. This

---

[151] Variable x that is introduced can be referred back to with two kinds of pronouns, x^{s}, a singular pronoun, and x^{n}, a numberless pronoun. They can be interpreted as follows:

$\|\|x^{s}\|\|^{g,h,w}$	=	$(g(x)(w)), if $x \in$ dom(g) & $(g(x)(w)) defined, else
	=	$(h(x)(w)), if $x \in$ dom(h) & $(h(x)(w)) defined,
	=	undefined otherwise
$\|\|x^{n}\|\|^{g,h,w}$	=	$g(x)(w)$, if $x \in$ dom(g), else
	=	$h(x)(w)$, if $x \in$ dom(h),
	=	undefined otherwise

[152] I only consider distributive readings of *they*.

[153] I assume that you can introduce the right predicates in CCT to make sense of $\#(x) > 1$ in the interpretation rules.

[154] If all noun phrase are analysed as anaphora, what happens to the claimed distinction in rigidity of proper names and definite descriptions? Kripke (1972) claimed that there exists an essential difference between proper names and definite descriptions. The former are rigid, their reference in modal contexts is fixed by the reference_context of the 'base' possibility, but the latter are not. However, it seems that once definite descriptions can be anaphorically used, this 'essential difference' has to be given up. The anaphoric variable can refer to a rigid set, and whether this set is a singleton set or not, the description will take over at least this much of the rigidity of the variable.

variable refers back to all individuals explicitly introduced in the discourse. To make sense of this, I interpret these terms as follows:

$||\alpha||^{g,h,w} = D(w)$

$||\delta||^{g,h,w} = \cup\{h(x)(w) | x \in dom(h) \& \$(h(x)(w) \text{ is defined}\}$

It seems reasonable to assume that anaphorically used definite descriptions normally pick up the set denoted by δ.

The suggestion that 'incomplete' definite descriptions only quantify over the explicitly introduced objects, the interpretation of δ, is in agreement with the claim made in Groenendijk *et al.* (1995b) that CCT is a natural framework for contextual restricted quantification. They claim that an incomplete description like *the man* refers to the unique man among the explicitly introduced objects. But this has an undesirable consequence; it is predicted that there exists an essential distinction between incomplete descriptions and pronouns. Incomplete descriptions should be treated as contextually restricted quantifiers, and pronouns as variables co-indexed with their antecedent. But I don't understand what the difference should be between the description *the male* and the pronoun *he*. Intuitively, I think, both are referentially used although the descriptive content of both noun phrases helps to determine, for each possibility of the context, to what objects they refer to. If it is assumed that incomplete descriptions and pronouns are referentially used, but that the descriptive content helps the hearer to determine what object the speaker had in mind, it is natural to take over Lewis's (1973) suggestion that a referentially used description of the form *the F* denotes *the unique most salient F* in the domain of discourse.[155] This means that in general, descriptions are anaphoric to the set denoted by the interpretation of α, the domain of discourse, and that the elements of this set should be ordered by salience. Salience is a property objects have in the worlds consistent with what is presupposed. That's why salience should not just be 'salience for the speaker'. That's also why the ordering of comparative salience can change so easily, and changes most systematically by conversation. How else to account for the inappropriateness of the following sequence: *The pig with floppy ears is grunting, but the pig is not grunting*, in a situation where the most salient pig for the speaker is not grunting? Although out of context, the speaker would talk about this most salient pig if he uses *the pig*, the use of *the pig* can no longer be appropriately used to refer to this pig after the first sentence is stated. If we give up the assumption made in original CCT that there exists an essential difference between salient objects and non-salient objects, the question arises what is left of the until now assumed specific property of indefinites, that they can make certain objects salient? Again, we should follow Lewis. By the use of an indefinite we do not transfer an object from the set of non-salient objects to the set of salient ones, but we just raise the salience of the object.

> I may say "A cat is on the lawn" under circumstances in which it is apparent to all parties to the conversation that there is some one particular cat that is responsible for the truth of what I say, and for my saying it. What I said was an existential quantification; hence, strictly speaking, it involves no reference to any particular cat. Nevertheless it raises the salience of the cat that made me say it. Hence this newly-most-salient cat may be denoted by brief definite descriptions or by pronouns, in subsequent dialogue. (Lewis, 1979b)

---

[155] Groenendijk *et al.* (1995b) seem to argue against this. In a sequence like *A man is walking in the park. Another man is walking in the park, too. The tallest man is whistling* the description *the tallest man* does not necessarily refer to the unique tallest man in the world, but just to the tallest man of the two man introduced earlier. But then, I think that a superlative like *the tallest man* should be analysed as something like 'the tallest individual of the set of most salient men'. This would work if as a result of the first two sentences there is no unique most salient man. And indeed, as noted by Groenendijk *et al.* (1995b) we cannot appropriately follow the first two sentences with a sentence like *The man is whistling*. According to my explanation this is so because the description *the man* does not pick out a unique most salient man in this conversational context. Note that we also cannot follow the first two sentences with *He is whistling*.

We have already explained how the apparent clash between the *existential* proposition expressed and the *specific* proposition meant can be resolved. What is relevant now is that we must implement the idea that by the use of an indefinite, we raise the salience of the object denoted by that indefinite noun-phrase. The problem that arises is how to do that in a framework closely related to original CCT?

Fortunately, this problem was solved for us by Krahmer (1995).[156] In terms of our framework, he proposed to extend the set of variables VAR with a function-variable *sw* that is interpreted with respects to an assignment *h* as $h(sw) \in \{N^{D(w)} | w \in W\}$. Thus, h(sw) denotes for each world *w* a total function from D(w) to the set of natural numbers. I will assume that if *d* and *d'* are elements of D(w), *d* will be more salient than *d'* in possibility <g,h,w> iff h(sw)(w)(d) < h(sw)(w)(d'). Now we assume that from the very beginning of a conversation, all assignments have *sw* in its domain. Because salience is not only due to conversation, the initial assignment functions that have only *sw* in its domain already give rise to non-trivial order relations. How do referring expressions raise the salience of certain objects, and how is salience relevant for the interpretation of definites? With respect to the first question I assume rather simplistically that the use of an indefinite makes the referent of this indefinite the most salient object in the domain of discourse.[157] Of course, in different worlds these referents of the indefinites might be different, that's one reason why they are different worlds. At the same time that a new discourse marker is introduced,[158] the salience of the referent of this discourse marker will become the highest. This can be implemented by changing the relativised definition of g[x]h with respect to the worlds in S as follows:

$$g[x]Sh \text{ iff } dom(h) = dom(g) \cup \{x\} \;\&\; \forall y \in dom(g)[(y \neq x \;\&\; y \neq sw) \rightarrow h(y) = g(y)] \;\&\; \forall w \in w(S)[h(sw)(w)(\$(h(x)(w))) = 1 \;\&\; \forall d \in D(w)[d \neq \$(h(x)(w)) \rightarrow h(sw)(w)(d) = g(sw)(w)(d) + 1]]$$

With respect to the second question, all we have to implement is that definites refer to the most salient object in the domain of discourse:

$$[[\iota^{y}_{x} A]](S) = \{<g[x/\hat{x}|A|^{g}_{h},y],h',w> | \exists h \subseteq h': <g,h,w> \in S \;\&\; \exists d \in \hat{x}|A|^{g}_{h},y(w) \;\&\; \forall d' \in \hat{x}|A|^{g}_{h},y(w)[d' \neq d \rightarrow h(sw)(w)(d') > h(sw)(w)(d)] \;\&\; h[x]Sh' \;\&\; h'(x)(w) = \{d\}\},$$

if $\forall \alpha \in S: y \in dom(\alpha) \;\&\; x \notin dom(\alpha)$, else undefined

where $\hat{x}|A|^{g}_{h},y$ denotes that function $f: W \rightarrow \wp(D)$ such that for any $w \in W$:

$$f(w) = \{d \in ||y||^{g,h,w} | [[A]](\{<g[x/\mathbf{d}],h,w>\}) \neq \emptyset\}$$

According to the above interpretation rule, only rigid individuals can be more or less salient with respect to each other, for properties there is still the absolute salient vs. non-salient distinction. To assume such an asymmetry is strange and should be given up. So, a singular pronoun like *he* either is referentially used and refers with respect to every world (or world-assignment pair) to the most salient male in that world, or it is used as a descriptive pronoun and refers in every world to that unique male individual that is the

---

[156] See also Von Heusinger (1995), but Krahmer's formalisation is richer.

[157] See Smaby (1979) for a more reasonable account of saliencenes. He argues, for instance, that for a subject predicate sentence where both subject and predicate contain anaphoric antecedents, the individual denoted by the subject expression will normally be more salient than the object denoted by the direct object.

[158] But once we have salience, discourse markers, or semantic variables, are of no central import anymore.

unique instantiation in that world of the most salient property in the relevant reference-context. To account for this in a formal way is certainly possible, but I won't do it here. It would only complicate the interpretation rules and gives no extra insights. All I want to do in this section is to pave the way for my later analysis of presuppositions in quantified contexts, and there I won't need a salience order of properties to make my point.

What will be crucial later is the insight that not only for singular definite noun phrases the salience order partly determined by the discourse is crucial. Noun phrases that are traditionally looked at as quantifiers behave in a similar way. We have already made some sense of this by claiming that also these quantifiers refer anaphorically back to a salient context set. But we have seen that definite descriptions do something more. They do not just anaphorically refer back to a salient individual, they are also able to *select* a particular individual among the salient ones. Because I have not yet treated presuppositions, until now this selection is done by the assertion made. But really, I think, we should select the relevant salient individual by presuppositional means. Just like presuppositions can be cancelled for reasons of informativety, a description like *the N* will probably not refer to what the hearer considers to be the most salient *N*, if the resulting assertion would be trivially true or false. And, just like hearers sometimes have to accommodate the context such that a presupposition made by the speaker is satisfied by the context, the hearer also sometimes has to accommodate the salience ranking such that the resulting assertion would be informative (cf. Lewis, 1979b). This, I would say, also holds for normal quantified expressions. If quantified sentences give rise to a presupposition, it is with the help of this presupposition that the appropriate salient context set is selected. To already make some sense of this I also introduce a special variable, β, and its interpretation $h(\beta)(w)$ will be $\wp(D(w))$. Also the elements of $\wp(D(w))$ are ordered with respect to their salience. In the beginning of the discourse the most salient element will probably be the domain of discourse, $\|\alpha\|^{g,h,w}$. Most arbitrary elements of $\wp(D(w))$ will be equally non-salient, and the elements that correspond with natural properties will be more salient than elements that do not. But, just like before, pointing and conversation can change this salience order. In chapter 4, I will argue that sentences like *Every man wears his badge*, and *Most men wear their badge* presuppose that every man has a badge. But this presupposition need not be a presupposition about the domain of discourse, it rather functions to select the maximal salient context set that satisfies this presupposition. If there is no such set, it will be predicted that the utterance made is odd. But that's for later.

I claimed that all noun phrases should really be thought of as being anaphoric in nature. I have shown already how to interpret definite noun phrases, but I haven't said anything yet about the way salience changes when an indefinite or a quantified noun phrase is interpreted. I propose to interpret them as follows (where $S[x:\approx d] = \{<g,h',w> \mid \exists h:<g,h,w> \in S \ \& \ h[x]Sh' \ \& \ \$(h'(x)(w)) = d\}$):

$$[[\exists^{y}{}_{x}(A)]](S) = \{<g'[^{x}/_{\hat{x}}|A|^{g}_{h},y],h',w> \mid <g,h,w> \in S \ \& \ g \subseteq g' \ \& \ h \subseteq h' \ \& \ \exists d \in \|y\|^{g,h,w} \ \& \ <g',h',w> \in [[A]](S[x:\approx d])\},$$
$$\text{if } \forall \beta \in S: y \in dom(\beta) \ \& \ x \notin dom(\beta), \text{ undefined otherwise}$$

$$[[D^{y}{}_{x}(A,B)]](S) = \{<g'[^{x}/_{\hat{x}}|A \wedge B|^{g}_{h},y],h,w> \mid [D]\,(\{d \in \|y\|^{g,h,w} \mid [[\hat{x}A]](\{<g,h,w>\})(d) \neq \emptyset\},$$
$$\{d \in \|y\|^{g,h,w} \mid \exists h':<g',h',w> \in [[\hat{x}(A \wedge B)]](\{<g,h,w>\})(d)\}) \ \& \ <g,h,w> \in S\},$$
$$\text{if } \forall \beta \in S: y \in dom(\beta) \ \& \ x \notin dom(\beta), \text{ undefined otherwise}^{159}$$

---

[159] Except for the introduction of properties and the anaphoric dependence, this interpretation rule is just the same as the one given in Kamp & Reyle (1993) and Dekker (1993). Kamp & Reyle introduce also a *set* by the interpretation of quantifiers. As I argued for in the main text, I prefer to introduce a non-rigid *property*. [D] is the usual set theoretical interpretation of the determiner D.

Thus, both indefinite and quantified noun phrases are treated as anaphoric quantifiers, and introduce properties, but only indefinite noun phrases introduce also rigid individuals. Note that because quantifiers introduce properties that can be anaphorically picked up by other quantifiers, and because I allow for the introduction of functional pronouns, I can easily account for sequences like *Every summer John rents a car to go to France. He usually takes it on the ferry*, where the second sentence intuitively means that for most summer events where John rents a car to go to France, he takes the car he rented on the ferry.

To make sense of the observation that definite incomplete descriptions like *the man* and *he* can select objects from a particular context set, it seems to me that all quantifiers, plural and singular ones, should really be interpreted with respect to $\|\beta\|^{g,h,w}$. The most salient element of this set is selected, but for singular definite noun phrases you also select, if possible, the most salient individual of this most salient set. I will leave it to the reader how this should be implemented, but the idea should be obvious.

If all these noun phrases are anaphoric and interpreted with respect to salience, what is left of Heim's novelty and familiarity condition? In a sense nothing changes. Formally, free variables should be defined, whereas introduced variables must be new. However, in this sense these constraints lose their interest once we work with salience. But the notions 'novelty' and 'familiarity' were intended to have something to do with the distinct presuppositional characters of the definite and indefinite noun phrases, too. The difference is that a definite noun phrase refers to the presupposed unique most salient object in its context set that satisfies the descriptive material, while this is not the case for indefinites. Like Van Deemter (1991), I would say that by means of scalar implicatures we can give a straightforward pragmatic explanation of this different behaviour of indefinite and definite noun phrases. If you use *a man* instead of *the man*, you strongly suggest that you cannot use *the man*, that the man you intend to refer to by the indefinite is not the only most salient man in the relevant context set for which the claim you make is true.[160]

2.10 CCT in possible worlds only

I claimed in the beginning of this chapter that CCT really made use of diagonalisation and that we can, if we want, account for presupposition states in terms of possible worlds only. We saw already how we can account in terms of possible worlds only to what object an indefinite refers when it is referentially used. What was not yet clear is how we can refer back to this individual with a pronoun or a brief definite description. Once the notion of salience is introduced we can make more sense of this. What is salient and what not for the participants in the conversation is a fact about the world. So, if the salience order of objects changes, the world changes too. Suppose that with the indefinite used in *An S is P* I have referred in v to a and in v' to b. If my claim is accepted, my presupposition state changes. From v we go to w, and from v' to w', where for instance w only differs from v with respect to the salience order. Thus, a will be the most salient object in w and b will be the most salient object in w' in the new presupposition state. If I use *the S* in a following sentence, it will refer to a in v, and to b in v'. I said that v only differs from v' with respect to their salience orders, but that is not quite true. The most important way in which v' differs from v, and the most interesting reason why it is a useful exercise to state CCT in possible worlds only, is that not only the salience order, but also the speakers presuppositions change when an assertion is accepted. If Presup(c,v) represents the speakers presuppositions of conversation c in v, the presuppositions of the speaker in w, Presup(c,w) will differ from the presuppositions of the speaker in v because in w the speaker assumes that the last assertion is accepted, which was not yet the case in v. If, given a speaker and a conversation, the speakers perspective, the salience order of the conversation, and the speakers presupposition are all functional dependent on the world,

---

160 One exception, we sometimes use an indefinite although we could use a definite, because we want to hide the identity of the individual we refer to (cf. Strawson, 1950).

we can state CCT in terms of possible worlds only. So let us say that s(c,w) is the salience order in w of conversation c, sp(c,w) the perspective of the speaker of c in w, and Presup(c,w) the speakers presuppositions of c in w.

In a possible world formulation of CCT, we must make a distinction between two kinds of facts, facts where the discourse is about, and facts about the discourse itself.[161] Traditionally, to individuate worlds only facts about the subject matter of discourse count. If we want to state CCT in possible worlds only, we have to individuate worlds in a more finer grained way. Still, we need to know when two worlds are equivalent with respect to the facts the discourse is about. Therefore we introduce equivalence classes of worlds. We say that w' $\in$ SSM(c,w) iff *w'* makes the same facts relative to the subject matter of conversation *c* true as *w* does.

In the following formulation of CCT, I ignore descriptive pronouns, assume for simplicity that predicates are at most 2-ary, and call the conversation *c*. If the presuppositions of the speaker are represented by an accessibility relation, also the presuppositions will change after the acceptance of an assertion. To avoid circularity, I assume that clauses that are not internally dynamic with respect to presupposition change have to undergo three kinds of interpretation rules, [[.]], [.] and {{.}}, where for any formula A, [[A]] and {{A}} are subsets of $\wp(W) \times \wp(W)$, and [A] is a subset of $\{<w,S> | w \in W \& S \in \wp(W)\} \times \{<w,S> | w \in W \& S \in \wp(W)\}$.

(1) $[[R(t_1,t_2)]](S) = \{w' \in W | <w', Presup(c,w')> \in [R(t_1,t_2)](\{<w,Presup(c,w)> | w \in S\})\}$

(1') $[R(t_1,t_2)](T) = \{<w',Presup(c,w')> | \exists <w,Presup(c,w)> \in T: w' \in \{\{R(t_1,t_2)\}\}(\{w\}) \& Presup(c,w') = \{\{R(t_1,t_2)\}\}(Presup(c,w))\}$

(1") $\{\{R(t_1,t_2)\}\}(S) = \{w' \in SSM(c,w) | w \in S \& \exists g,s,\phi,\phi': g = Term(R(t_1,t_2), <\phi',\varnothing,s(c,w),w>) \& <[[t_1]]^{\phi',g,s(c,w),w}, [[t_2]]^{\phi',g,s(w),w}> \in I_w(R) \& sp(c,w)) = \phi \& s(c,w')([[t_1]]^{\phi,\varnothing,s(c,w),w}) = 1 \& s(c,w')([[t_2]]^{\phi,Term(t_1(<\phi,\varnothing,s(c,w),w>),s(c,w),w}) = 2 \& \forall d \in D(w)[d \notin \{[[t_1]]^{\phi',g,s(c,w),w}, [[t_2]]^{\phi',g,s(w),w}\} \Rightarrow s(c,w')(d) = s(c,w)(d) + 2 \& sp(c,w') = \phi$ except that for all $x \in VAR$ and for all i: $1 \le i \le 2$: if $t_i = \varepsilon x P$, then $\phi'(I_W(P)) = \phi(I_W(P) - \{\phi(I_W(P))\})\}$[162]

In this rule, the definition of the predicate 'Term' is as in section 4 of this chapter, and terms are evaluated as follows:

---

[161] There is of course no hard distinction possible; sometimes discourses can be about the words used in the discourse itself.

[162] In the above interpretation rules we have assumed that from acceptance of a sentence we go distributively from one world to another. But if we do everything in terms of worlds, we have to assure that there are enough worlds. To assure this we assume that every model <W,I> for our language obey the following constraints:

For every w $\in$ W, and A $\in$ $FORM_L$: $\exists w' \in SSM(c,w)$: $Presup(c,w') = [[A]](Presup(c,w))$;

For every w $\in$ W, and A $\in$ $FORM_L$: $\exists w' \in SSM(c,w)$:

$s(w')(\{d \in D | [[A]]^{\phi,g[x/d],s,w} = 1\}) = 1$ and $\forall d \in D(w)$: $s(w')(d) = s(w)(d) + 1$;

For every w $\in$ W, g $\in$ G, and A $\in$ $FORM_L$: $\exists w' \in SSM(w)$: $sp(w')) (= \phi') = sp(w)) (= \phi)$ except that

$\phi'(\{d \in D | [[A]]^{\phi',g,s(w),w} = 1\}) = \phi(\{d \in D | [[A]]^{\phi',g,s(w),w} = 1\} - \{\phi(\{d \in D | [[A]]^{\phi',g,s(w),w} = 1\})\})$.

$[[t]]^{\phi,g,s,w}$ = g(t), if t is a variable,
= I(t)(w), if t is an individual constant,
= $\phi(\{d \in D | [[A]]^{\phi,g[x/d],s,w} = 1\})$, if $t = \varepsilon_x A$,
= $s(\{d \in D | [[A]]^{\phi,g[x/d],s,w} = 1\})$, if $t = \iota_x A$, and
= * otherwise

Non-atomic clauses are interpreted in the following way:

(2) $[[\sim A]](S) = \{w \in S | \sim\exists w' \in SSM(c,w): w' \in [[A]](\{v \in SSM(c,w") | w" \in S \ \&\ s(c,v) = s(c,w") \ \&\ Presup(c,v) = Presup(c,w")\})\}$

(3) $[[A \wedge B]](S) = [[B]]([[A]](S))$[163]

(4) $[[\forall xA]](S) = \{w' \in W | \langle w', Presup(c,w')\rangle \in [\forall xA](\{\langle w, Presup(c,w)\rangle | w \in S\})\}$

(4') $[\forall xA](T) = \{\langle w', Presup(w')\rangle \in T | w' \in \{\{\forall xA\}\}(\{w\}) \ \&\ Presup(c,w') = \{\{\forall xA\}\}(Presup(c,w))\}$

(4") $\{\{\forall xA\}\}(S) = \{w \in S | \forall d \in D(w), \forall\phi: [[A]]^{\phi,\{\langle x,d\rangle\},s(w),w} = 1\}$[164]

Given this presuppositional accessibility relation, we can now interpret epistemic *might* constructions in a distributive way:

(5) $[[\Diamond A]](S) = \{w \in S | [[A]](Presup(c,w)) \neq \varnothing\}$

Note that in this way $[[\Diamond A]](S)$ need not be an all or nothing affair anymore. The context consists of those possibilities of which the hearer thinks that they might be the actual world. But the hearer doesn't know exactly what the speaker is presupposing, and thus there might be two worlds in S, *w* and *w'*, such that $Presup(c,w) \neq Presup(c,w')$. In particular, it might be that A is consistent with what is presupposed by the speaker in *w*, but not with what the speaker presupposes in *w'*. As a result, the speakers claim that it might be the case that A can be informative for the hearer. I will argue in chapter 4 that it is useful to assume that it is unclear to the hearer what is presupposed by the speaker to account for presupposition accommodation.

[163] Thus, no free variables in B can be bound by quantifiers in A.

[164] Where $[[A]]^{\phi,g,s(w),w}$ is determined in a similar way as $[[A]]^{\phi,g,w}$ was determined in §2.6.

Chapter 3

Anaphoric relations across attitude contexts

According to the received view in semantics, so-called unbound pronouns - pronouns not bound by a quantifier Q inside the smallest clause containing Q - should either be treated as abbreviations for the antecedent clause or as variables bound by a dynamic existential quantifier. Geach's notorious Hob-Nob sentences, exemplifying intentional identity attributions, have always been a threat to this assumption. In a recent series of insightful papers (Edelberg 1986, 1992, 1995), Edelberg has also challenged the traditional *realist* conception of semantics, according to which sentences denote propositions that are true or false in the actual world. He argues that if we look at both *de re* belief attributions and at attitude attributions of intentional identity, a *perspectivalist* semantic theory will be more economical than a realist one. According to Edelberg's perspectivalist semantic theory, what is expressed by a sentence is not said to be true or false in a world, but true or false relative to a theory or belief state. In this chapter I will discuss how successfully the problems can be handled in terms of the traditional assumptions. I will argue that (i) a realist account of *de re* attributions and Hob-Nob sentences need not really be more complex than the perspectival account proposed by Edelberg, but also that (ii) the realistic and externalistic dimension is crucial for a proper account of intentional identity attributions. On the other hand, I will suggest that the phenomenon does show the limits of the received doctrine with respect to the analysis of so-called unbound pronouns, if Edelberg's (1986) asymmetry problem has to be accounted for within semantics.

3.1 The problem of intentional identity

Problematic for all semantic accounts of anaphora is that a pronoun occurring in the embedded clause of an attitude attribution can have as its syntactic antecedent an indefinite in the embedded clause of an earlier attitude attribution. In a logical language this is not difficult to represent if the indefinite is interpreted *de re*. But the problem is that this doesn't always seem to be the case. This is the problem that was discussed under the heading of *intentional identity* by Geach (1967), and called the problem of *de dicto pronouns* by those who were working in the tradition of Montague semantics. Examples of these sentences include the following:

(1) John believes that *a woman* broke into his apartment.
He believes that *she* is now hiding from the police.
(2) Carl wants to catch *a fish* today, and he wants to eat *it* afterwards.
(3) Hob believes that *a witch* blighted Bob's mare,
and Nob believes *she* killed Cob's sow.

On the intended readings of these sentences, the attitude attributions can be true without there being a woman about which John has the relevant beliefs, a fish the desire attributions to Carl are about, or a witch that is responsible for the beliefs of Hob and Nob. For (3), there does not even have to be an existing individual that is the focus of both Hob's and Nob's beliefs. This is shown by the following Geachian story:

> Last night, Bob's mare became quite ill. Hob, who tends Bob's barn, inferred that a witch blighted her. This morning Hob said to his friend, Nob, "A witch blighted Bob's mare." Nob believes what Hob has told him. He thinks for a moment, and says, "Cob's sow died early this morning. I'll bet the same witch killed the sow, too." But in fact both animals fell ill due to perfectly natural causes. (Edelberg, 1986, pp. 1-2)

According to this story, (3) would be true. In the Geachian tradition, anaphoric elements are treated as bound variables; but the problem is that there is no way to bind the variable that represents the pronoun in the second clause by the quantifier that represents the indefinite in the first clause if you can quantify only over existing individuals. In the framework of traditional Montague semantics, the following translations might be tried:

(3a) $Bel(h, \exists x[witch(x) \wedge BBM(x)]) \wedge Bel(n, KCS(x))$

(3b) $Bel(h, \exists x[witch(x) \wedge BBM(x) \wedge Bel(n, KCS(x))])$

(3c) $\exists x[witch(x) \wedge Bel(h, BBM(x)) \wedge Bel(n, KCS(x))]$

If pronouns are treated as bound variables, it seems that the only possible way to go is to use either representation (3b) or (3c). Unfortunately, (3b) doesn't give the intended reading because the attitude attribution doesn't seem to say anything about what Hob believes about Nob beliefs, and representation (3c) does not predict (3) to be true in the above story because in fact witches do not exist.

One might think that the problem can be solved if the pronoun used in the second conjunct of (3) should be read as a *descriptive pronoun*. Moreover, we have seen in the last chapter that descriptive pronouns can be implemented into CCT such that they don't obey anaphoric island constraints. It might be proposed that (3) should be represented by (3a) after all, and that the pronoun *she* is an abbreviation for *the witch who blighted Bob's mare*. Alternatively, it might be proposed that the pronoun is an abbreviation for *the witch Hob thinks blighted Bob's mare* or *the object of which Hob thinks it is a witch that blighted Bob's mare*. Unfortunately, Geach already showed that none of those suggestions can solve the problem. Consider the following story:

The Gotham city newspapers have reported that a witch, referred to as "Samantha", has been on quite a rampage. According to the article she has been blighting farm animals and crops and throwing people down wells. In reality, there is no such person: the animals and crops all died of natural causes, and the people found at the well-bottoms had all stumbled in by accident in a drunken stupor. The news reporters simply assumed that a witch was responsible for all the mishaps, and dubbed her "Samantha". Hob and Nob both read the *Gotham Star* and, like most folks, they believe the stories about the witch. Hob thinks Samantha must have blighted Bob's mare, which took ill yesterday. Nob thinks Samantha killed his friend Cob's sow. Nob has no beliefs at all about Hob or about Bob's mare; he is unaware of the existence of either. (Edelberg, 1986, p. 2)

Although in this situation (2) has a reading that is true, it is clear that in this situation none of the proposed descriptions can be an abbreviation for the pronoun in the second conjunct of (3).[165]

From these problems some have concluded that variable *x* should really range over *non-existent* objects, and that cases of intentional identity should be translated as in (3c) after all. In cases of intentional identity, a *de re* belief attribution is made about a specific object that might be non-existent. The problem with this assumption is that a sentence like (2) doesn't seem to be about a specific (maybe non-existent) fish at all. There does not need to be one specific fish John's belief is about such that John believes he will catch it and wants to eat it afterwards to make the attitude attribution true[166]. Let's call this problem *the specificity problem*. Moreover, for (3) for instance it should be predicted that in all of Hob's belief alternatives there is a witch who blighted Bob's mare, something that is not guaranteed if we represent (3) by (3c). These two problems suggest that we should represent intentional identity attributions in a non-Montagovian way, as in (3d):

(3d) $\exists xBel(h,W(x) \wedge BBM(x)) \wedge Bel(n, KCS(x))$

In fact, Slater's proposal (1988) boils down to this. According to it, Hob and Nob have a belief about a specific object, but all we know about this object is that Hob thinks that *it* is a witch that blighted Bob's mare, and Nob believes that *it* killed Cob's sow. But intuitively (3) can be true, without any specific object satisfying the above conditions. The reason is

---

[165] The only possibility consistent with the story is to assume that the descriptive pronoun takes wide scope over the belief predicate, and is an abbreviation for *the witch of which Hob thinks it blighted Bob's mare*. I will argue later, following Buridan, that this gives rise to an implausible analysis.

[166] See also Haas-Spohn (1986).

that there need not be one actually existing object that is responsible for the relevant beliefs of Hob and Nob. Hob believes none of the individuals he has ever come across to be a witch; thus none of them satisfies the property expressed by $\hat{x}\,\mathrm{Bel}(h,W(x) \wedge \mathrm{BBM}(x))$ (cf. Buridan, 1350). Arguing that variables should also range over non-existing objects does not help, if it is assumed that indefinites occurring in embedded sentences of belief attributions will be represented by a formula where the corresponding existential quantifier has wide scope with respect to the belief predicate. That would give rise to the unwanted prediction that for the first conjunct of (2) to be true, there must be a specific object of which Carl believes that *it* is a fish that he will catch today.

All these problems suggest that we should indeed represent a sentence like (3) by (3d), but that the variables should not range over specific objects, but over *individual concepts*, instead. Something like this was proposed by Saarinen (1978) to account for intentional identity attributions.[167] He assumed that variables range over individual concepts, and that these concepts don't have to be instantiated in the actual world. However, this suggestion is problematic for two reasons. First, as we have seen in chapter 1, if variables always range over all individual concepts, *de re* belief attributions are predicted to be too easily true. Second, if we don't restrict the range of the variables, by Saarinen's proposal we would predict that attributions of the form (4a) are equivalent with attributions of the form (4b):

(4a) $\exists x\mathrm{Bel}(a,Px) \wedge \mathrm{Bel}(b,Qx)$

(4b) $\exists x\mathrm{Bel}(b,Qx) \wedge \mathrm{Bel}(a,Px)$

However, Edelberg (1986, 1992, 1995) observed that intentional identity attributions are in general not symmetric. Consider the following case:

> Arsky and Barsky investigate the apparent murder of Smith, and they conclude that Smith was murdered by a single person, though they have no one in mind as a suspect. A few days later, they investigate the apparent murder of a second person, Jones, and again they conclude that Jones was murdered by a single person. At this point, however, a disagreement between the two detectives arises. Arsky thinks that the two murderers are completely unrelated, and that the person who murdered Smith, but not the one who murdered Jones, is still in Chicago. Barsky, however, thinks that one and the same person murdered both Smith and Jones. However, neither Smith nor Jones was really murdered. (Edelberg, 1995, p. 317)

For this case we intuitively find (5) acceptable, but not (6):

(5) Arsky believes that someone murdered Smith, and
Barsky believes he murdered Jones.
(6) Barsky believes that someone murdered Jones, and
Arsky believes he murdered Smith.

Intentional identity attributions are in general *not symmetric*, although Saarinen's proposal wrongly predicts them to be. Edelberg called this problem the *asymmetry problem about intentional identity*. Note, too, that any proposal that seeks to account for intentional identity by representing sentences like (3) by (3d) and by allowing for quantification over non-existing objects fails to explain this asymmetry.[168]

A different but related problem is discussed by Edelberg under the heading of *the variable aboutness problem of attitudes de re*. The problem is related to the following case:

> Smith and Jones are dead. A single person murdered both of them. Detective Arsky investigates both cases, and comes to believe that someone murdered Smith and that someone murdered Jones, but he doesn't have

[167] See also Zeevat (1996).

[168] The same holds for the proposal of footnote 1 to explain intentional identity attributions by descriptive pronouns having wide scope over the belief operator.

anyone in particular in mind as a suspect. Arsky does not believe that Smith's murderer and Jones's murderer are the same person. (Edelberg, 1995, p. 318)

The problem is to account for the intuition that on their most straightforward readings, (7) is true, while (8) is false:

(7) Someone murdered Smith, and Arsky thinks he didn't murder Jones.[169]
(8) Someone murdered Smith, and Arsky thinks he murdered Jones.

The problem for the approach where variables range over concepts is that it is predicted that (8) as well as (7) is true, because there is a single concept, *the murderer of Jones*, whose instantiation in the actual world murdered Smith and whose instantiation in Arksy's belief worlds also murdered Jones in each of them.

In the last chapter we argued that pronouns are either referential or descriptive. One might think that in the above cases the asymmetry problem can be solved if the pronouns in the second sentences of (5) - (8) are treated as descriptive pronouns. But Edelberg gives a counterexample to this proposal, too:

Monday: Smith and Jones have been shot, at opposite ends of Chicago. Arsky and Barsky are investigating both cases, but neither knows that Smith is the mayor or that Jones is the commissioner. Smith and Jones, though hospitalized, are (and are known by both detectives to be) still alive. Arsky and Barsky have discussed the two cases at length, and though they think someone shot Smith and that someone shot Jones, both believe the two cases are entirely unconnected. At this time, neither has anyone in mind as a suspect. *Tuesday*: Both Smith and Jones have died of their gunshot wounds. Arsky knows Smith died, and thus now believes that the person who shot Smith murdered him, but doesn't know Jones is dead. Likewise, Barsky knows Jones died, and thus now believes that the person who shot Jones murdered him, but doesn't know Smith is dead. After reflecting on certain similarities between the two cases, Barsky infers that the man who shot Smith is the same person as the man who shot Jones. He communicates this to Arsky, saying, "The man who shot Smith is the man who shot Jones." Arsky disagrees, but Barsky persists in his opinion. (Edelberg, 1986, pp. 16-17)

On Tuesday, (5) is true and (6) is false on their most natural readings. However, this asymmetry cannot be explained by treating the pronouns as abbreviations for descriptions recoverable from the clause in which the indefinite occurs. Barsky does not believe that Smith was murdered, and Arsky does not believe that Jones was murdered[170].

We have three kinds of problems now. First, we have cases like (1) and (2), where only one agent is involved and the pronoun in the second sentence does not refer back to a specific existing object that the speaker refers to. Second, we have *de re* attributions like (7) and (8), where the pronoun in the second sentence *does* refer back to such a specific existing objects. And third, we have intentional identity attributions like (3), where two agents are involved and the pronoun does not refer, for the speaker, to a specific existing individual. For *de re* attributions we have to account for the variable aboutness problem; and for intentional identity attributions with more agents involved we have to account for the asymmetry problem.

The Context Change Theory discussed in the last chapter is a framework that can handle anaphoric dependencies across sentential boundaries. It is only to be expected that the intentional identity cases discussed above could be handled in this framework, too. Indeed, this is what I believe. But as we will see, this is not as straightforward as one might hope.

In dynamic theories of meaning the meaning of a sentence A, [[A]] , is a function from contexts to contexts. If a sentence is accepted, the new context consists only of possibilities that make the sentence true. This is also the case for a belief attribution like *John believes*

[169] I am making use of this rather awkward phrasing to keep scope matters clear.
[170] Note that this example also shows that the asymmetry problem is not solved by giving up the assumption that the connective '∧' is symmetric.

that A. For the propositional case, a context, *I*, is represented by a set of possible worlds, and [[A]] (I) is the set of possible worlds in *I* that make A true. The belief sentence above would be true in world *w* iff A is true in all worlds *w'* that are doxastically accessible to John, K(j,w). Heim (1992) has given the following interpretation rule for belief sentences in a dynamic framework:

$[[\text{Bel}(j, A)]](I) = \{w \in I \mid K(j, w) = [[A]](K(j, w))\}$[171]

If a world *w* of *I* is such that A is not true in all worlds of K(j,w), it is eliminated from the context. 'K(j,w)' fulfils two roles in Heim's interpretation rule. First, it represents what John believes in *w* as a *fact* in the model. Second, it functions as the *context of interpretation* for embedded sentences of belief and other attitude attributions.[172] The second role is important for the analysis of anaphora and presuppositions. For instance in Heim (1992), if A triggers presupposition B, Bel(j, A) is defined only for those worlds in which the set of doxastically accessible worlds for John entails the presupposition. Bel(j, A) is said to be unacceptable in context *I* if there is a world *w* in *I* such that K(j,w) $\not\subseteq$ B. It follows that for Bel(j, A) to be acceptable in *I*, it should be the case that $\cup\{K(j, w): w \in I\} \subseteq B$.[173]

In Heim (1992) not much attention is paid to the problem how to account foranaphoric dependencies across attitude reports. The anaphoric relation between the pronoun *she* in the second sentence of (1) repeated here,

(1) John believes that *a woman* broke into his apartment.
He believes that *she* is now hiding from the police.

and its antecedent *a woman* of the first sentence cannot be handled properly. The context of interpretation for the embedded clause of the second sentence does not contain enough information to interpret the pronoun. CCT was developed to account for anaphoric relations across sentential boundaries. In CCT, information states are represented by sets of world-assignment pairs. The assignments are *partial* functions representing the information that in the discourse we have talked about only a limited set of objects.

It seems that the main idea underlying the CCT approach can be maintained if for the analysis of indefinites we use *guarded assignments*. The guarded assignment, x:=*, extends the domain of the assignment function with *x*, if *x* is not yet in its domain, and does nothing otherwise. Where usually a sentence like *A man is sick* is translated as something like $\exists x Px$, we can translate it simply as Px, and we interpret $Px_1,...,x_n$ in context S as follows:

$[[Px_1,...,x_n]](S) = [[x_1:=* \wedge ... \wedge x_n:=* \wedge \bullet Px_1,...,x_n?]](S)$,

where the clause 'x:=*' is interpreted as suggested above, and

---

[171] Heim attributes the formulation of the rule to Kamp.

[172] Stalnaker (1988) and Zeevat (1992) give the following kind of interpretation rule:
$[[\text{Bel}(j, A)]](I) = \{w \in I \mid K(j, w) \subseteq [[A]](\cup\{K(j, w): w \in I\})\}$
In this way, the second role is taken over by that what *we* presuppose is compatible with what John believes, $\cup\{K(j, w): w \in I\}$.

[173] Zeevat's interpretation rule will have the same effect for the main level when the embedded sentence denotes a persistent proposition. There will be a difference, however, when Veltman's (1990) epistemic *might* is introduced. By Zeevat's interpretation rule, Bel(x, $\Diamond$A) functions as a global test, just like the presupposition trigger discussed in the main text, but this is not the case for Heim's pointwise interpretation rule.

$[[\bullet P x_1,\ldots,x_n?]](S) = \{\langle g,w\rangle \in S \mid \langle g(x_1),\ldots,g(x_n)\rangle \in I_w(P)\}$.

If we make these assumptions, it seems natural to interpret belief sentences as follows:

$[[Bel(x, A)]](S) = \{\langle g,w\rangle \in S \mid K(g(x), w) = [[A]](K(g(x), w))\}$

where also K(g(x),w) is an information state and thus modelled by a set of world-assignment pairs.

But in this way we cannot account any more for the intuition that only a few 'belief objects' were introduced in the discourse. The asymmetry of the use of pronouns cannot be accounted for. The novelty and familiarity condition don't make sense anymore if we assume that the update of K(g(x),w) by A does not change K(g(x),w). To account for novelty and familiarity in belief contexts it seems that although A might be true in the worlds of K(g(x),w), the indefinites occurring in A should still enrich the assignment functions of this K(g(x),w). The problem is that there is no way to represent this information growth of K(g(x),w) in the corresponding possibility of the main context. World *w* is not eliminated, and assignment *g* does not grow. Let's call this problem the novelty/familiarity problem of intentional identity for CCT.

In the rest of this chapter I will discuss various approaches that can solve the novelty/familiarity problem of intentional identity for CCT. In the first approach, what I call *constructive update semantics*, I try to solve the novelty/familiarity problem in maybe the most immediate way. I will claim that the resulting framework can technically solve our problem, but that this is done in a conceptually wrong way. The second approach, what I call the *modal subordination account*, takes more serious the idea that it is *the speaker* who is responsible for the use of indefinites in embedded clauses of belief attributions. However, this approach cannot by itself account for Hob-Nob sentences, and, as I will argue, gives too weak readings to belief sentences in which an anaphor is used in the embedded clause picking up an indefinite used in another belief attribution. Afterwards, I discuss a method to discuss Hob-Nob sentences in terms of descriptive pronouns. The approach can handle the phenomena, but at the cost of making the representation of the attributions very context dependent. Then, I try to account for the phenomena on the basis of the received view with respect to unbound anaphora by letting the variables range over *belief objects*. I will look at some different ways this idea can be implemented, but I will conclude that the asymmetry problem cannot be solved if anaphora are still treated as bound variables bound by a dynamic existential quantifier. Finally I sketch how the asymmetry between the belief attributions can be accounted for when pronouns are treated as referential expressions, as argued for in chapter 2.

3.2 Constructive update semantics

The first possible way to solve the novelty/familiarity problem is to argue that the complications arise because an information state is modelled by a set of *total* worlds and *partial* assignments. This forces one to make updates by belief-sentences *eliminative*, but then one cannot account for the fact that the derived context(s) should grow (the assignment functions at least). For this reason it seems nice to make updates never (or almost never) eliminative. It can be argued that it is better and more intuitive to represent an information state by a set of *partial* worlds and *partial* assignments. How should one represent a partial world? From an information theoretic perspective it seems to be a good idea to take over Van Fraassen's (1966) idea of *supervaluation* by representing a partial world as a situation (valuation function) that can *grow* into a set of total worlds (its supervaluations). In other words, this situation *stands for* the set of its supervaluations. The supervaluations (total worlds) are just classical valuation functions.

In this section I will proceed as follows: First I define a model with partial situations and put some constraints on it. Then I give classical truth conditions for clauses on the maximal situations. In terms of that we define via supervaluation the truth and falsity conditions for clauses on the other situations. Then we can define a selection-function, and in terms of that, the update function. Finally I will discuss how this system could be extended in order to account for belief attributions.

I will use the set G of partial assignments partially ordered by the subset relation, and the set S of partial possible worlds partially ordered by $\leq$. A state is now an element of $S \times G$. A frame for our language is now a triple $<S, \leq, D>$ where $<S, \leq>$ is a partial order, and D is the domain. Let us assume that every situation has the same domain D. We will call the maximal elements in $<S, \leq>$ *worlds*, thus w is a world iff for all s', if $w \leq s'$, then $w = s'$. We assume that every situation can grow into a world:
Completability: $\forall s \in S: \exists w \in S: s \leq w$.

A model M for our language is now a triple $<F, I^+, I^->$, where $F = <S, \leq, D>$ and the following conditions obtain for all s, s' $\in$ S, and all n-place predicates P:

1. $I^+_S(P) \subseteq D^n$ and $I^-_S(P) \subseteq D^n$
2. consistency: $I^+_S(P) \cap I^-_S(P) = \emptyset$,
3. monotonicity: $s \leq s' \Leftrightarrow I^+_S(P) \subseteq I^+_{S'}(P)$ and $I^-_S(P) \subseteq I^-_{S'}(P)$,[174]
4. totality: for all w: $I^+_W(P) \cup I^-_W(P) = D^n$
5. uniqueness: for all $s \in S$, and all $<d_1,...,d_n> \in D^n$:
   if $<d_1,...,d_n>$ is in the gap of $I_S(P)$, then
   $(\exists s'>s: <d_1,...,d_n> \in I^+_{s'}(P)$ and $\forall t>s[<d_1,...,d_n> \in I^+_t(P) \Rightarrow s' \leq t])$ and
   $\exists s''>s: <d_1,...,d_n> \in I^-_{s''}(P)$ and $\forall t>s[<d_1,...,d_n> \in I^-_t(P) \Rightarrow s'' \leq t]$)

Local truth- and falsity-conditions for atomic formulae in situations can now be given in terms of truth and falsity of these formulae in the worlds extending that situation, with the help of supervaluation.

$<g(x_1),..., g(x_n)> \in I^+_s(P)$ iff for all $w \geq s$: $<g(x_1),..., g(x_n)> \in I^+_w(P)$
$<g(x_1),..., g(x_n)> \in I^-_s(P)$ iff for all $w \geq s$: $<g(x_1),..., g(x_n)> \in I^-_w(P)$
iff (by totality) for all $w \geq s$: $<g(x_1),..., g(x_n)> \notin I^+_w(P)$[175]

In order to define the selection function and the update function, we first have to introduce the following abbreviations:

$g[^x/_d] = \{<y,g(y)> | y \in dom(g) \text{ and } y \neq x\} \cup \{<x,d>\}$
$g[x]h$ iff $h = g[^x/_{h(x)}]$
$<g,s> \leq <h,s'> := s \leq s' \ \& \ g \subseteq h$

---

[174] This condition just means that the order relation between the situations is definable in terms of the interpretation functions in the model.

[175] Let Δ be the set of the definable propositions in **L**. The following equivalence then follows from our definitions (given a classical definition of $w \models A$):
$s \leq s' \Leftrightarrow \forall A \in \Delta: \{w: s \leq w\} \subseteq \{w: w \models A\} \Rightarrow \{w: s' \leq w\} \subseteq \{w: w \models A\}$.
It only differs from normal possible world semantics if in the latter there are two worlds that are indistinguishable as far as the definable propositions are concerned.

$S \leq S'$ iff $\forall \alpha' \in S' : \exists \alpha \in S : \alpha \leq \alpha'$

By making the uniqueness assumption we can recursively define the following selection-function f for an arbitrary possibility <g,s> (Of course, f is not functional, in that it takes a situation assignment pair to another situation assignment pair. As usual, disjunctions and existential quantification make that impossible) :

$f_s(Px_1,\ldots, x_n, g) = \{<g,s'> | s \leq s' \ \& <g(x_1),\ldots, g(x_n)> \in I^+_{s'}(P) \ \&$
$\forall t > s[<d_1,\ldots,d_n> \in I^+_t(P) \Rightarrow s' \leq t]\}$,
if $\forall x_i : 1 \leq i \leq n$: $g(x_i)$ is defined, else undefined

$f_s(x_1 = x_2, g) = \{<g,s> | g(x_1) = g(x_2)\}$,
if $\forall x_i : 1 \leq i \leq 2$: $g(x_i)$ is defined, else undefined

$f_s(\sim A, g) = \{<g,s'> | s \leq s' \ \& \ f_{s'}(A,g) = \emptyset \ \& \ \forall t > s: f_t(A, g) = \emptyset \Rightarrow s' \leq t\}$

$f_s(A \wedge B, g) = \cup \{f_{s'}(B, h) | <s',h> \in f_s(A, g)\}$

$f_s(\exists xA, g) = \cup \{f_s(A, g[^x/_d]) | d \in D\}$, if $x \notin$ dom(g), undefined otherwise

The formulae $A \vee B$, $A \rightarrow B$ and $\forall xA$ are abbreviations of respectively $\sim(\sim A \wedge \sim B)$, $\sim(A \wedge \sim B)$ and $\sim\exists x \sim A$.

In terms of the above selection-function, we can now define the update function of information states [A] $[\subseteq \wp(G \times S) \times \wp(G \times S)]$ of formuale A of **L**:

$[A](X) = \cup \{f_s(A, g) | <g,s> \in X\}$

A formula A is *acceptable in X*, $X \Vdash_d A$ iff for all $<g,s> \in X$: it is the case that for all w that extend s, $w \geq s$, A is true in w with respect to g: $f_w(A, g) \neq \emptyset$. A formula A is *accepted in X*, $X \Vdash_S A$, iff $X = [A](X)$. Finally, B is *entailed* by A, $A \Vdash_{d/S} B$, iff for all contexts X: $[A](X) \Vdash_{d/S} B$.

If you learn in <s,g> the formula $\sim\exists xPx$, then s' is a minimal extension with respect to $\sim\exists xPx$ iff $I_{s'^-}(P) = D$. This can be shown as follows:

1	$<g,s'> \in f_s(\sim\exists xPx,g)$ iff	
2	$f_{s'}(\exists xPx,g) = \emptyset$ iff	(clause for negation)
3	$\forall d \in D: f_{s'}(Px, g[x/d]) = \emptyset$ iff	(clause for $\exists xA$)
4	$\forall d \in D: \sim\exists w \geq s': d \in I^+_w(P)$ iff	(clause for atomic formulae)
5	$\forall d \in D: d \in I^-_{s'}(P)$ iff	(consistency and superfalsity)
6	$I_{s'^-}(P) = D$	(definition of $I_s^-(P)$)

It can be proved that constructive update semantics gives rise to the same consequence relation as original CCT. Thus, if $W(X) = \{<g,w> | \exists s: <g,s> \in X \ \& \ s \leq w\}$, then it can be proved that $X \Vdash_d A$ iff $W(X) \models_d A$. This is shown in the appendix of this chapter.

Update semantics has only two kinds of updates: either you enrich the assignment-function, or you eliminate worlds. We have three kinds of updates: the assignment-function is enriched, we eliminate possibilities (for equality, or when it is already presupposed that

what is expressed by a sentence is false), or the situation is getting richer. We will suggest that for this reason we can handle belief-clauses in such a way that technically we can give sense to the notions of 'novelty' and 'familiarity' also with respect to these clauses.

Let us now add the predicate Bel to the language and add to the model a family of accessibility relations $\{K_d \subseteq S \times \wp(S \times G) | d \in D\}$. The goal is to go the same way for belief as we did for normal predicates. Sentence A is believed by *x* in *s* iff A is believed by *x* in all worlds that extend *s*.

How to do that? What is believed by *x* in a world is represented by a set of possibilities. Normally these possibilities are simply possible worlds. However, it might be assumed that to account for anaphoric dependencies across attitude contexts we should represent such a belief state by a set of world-assignment pairs.

So let us assume that for each agent *a* and for each world *w*, K(a,w) consists of a set of world-assignment pairs. Let us also assume that all elements of K(a,w) have the same domain: $\forall <w'g>,<w'',h> \in K(a,w)$: dom(g) = dom(h). If we define the domain of K(a,w), D(K(a,w)), to be $\cap\{dom(g) | \exists w': <w',g> \in K(a,w)\}$, it follows that D(K(a,w)) is well defined. Now we want to determine K(a,s) for situations that are smaller than worlds. I will define them in the following way:

$$K(a,s) := \{<g,s'> | dom(g) = \cap\{D(K(a,w)) | w \geq s\} \text{ and } \forall w \geq s: \{<g,s'>\} \leq K(a,w) \text{ and } \forall t: \forall w \geq s: \{<g,t>\} \leq K(a,w) \Rightarrow t \leq s'\}.$$

If we would then re-define when $s \leq s'$ in such a way that also belief states are taken into account, we can state the selection function applied to a belief sentence as follows:

$$f_s(Bel(x, A), g) = \{<g,s'> | s \leq s' \;\&\; K(g(x), s') \geq [A](K(g(x), s)) \;\&\; \forall t > s[\, K(g(x), t) \geq [A](K(g(x), s)) \Rightarrow s' \leq t]\}$$

So, if an information state is updated with a belief sentence, the main context is updated by really updating the belief context(s), too. Just like normal updating, a belief context grows by either eliminating worlds, by enlarging the domain of the assignment function(s), or by enlarging the component situations.

In the most minimal situation, s_0, we can assume that $\cap\{D(K(a,w)) | w \geq s_0\} = \varnothing$. This changes once we use a belief sentence with an indefinite occurring in the embedded clause. Suppose we have only made the following claim:

(9) John believes that a_x man is walking in the park.

Then the main context will only consist of situations *s* such that for every element <s',g> of K(j,s), dom(g) = {x} and g(x) walks in the park in every world of *s'*, and there is (probably) no situation *s''* such that g(x) walks in the park in every world of *s''* and s'' > s'. Thus, if X is the context, if we start talking about John's belief, $\cap\{D(K(a,w)) | \exists <g,s> \in X: w \geq s\}$ is getting larger, while $\cup\{W(K(a,s)) | \exists <g,s> \in X\}$[176] is getting smaller. The definedness conditions for indefinites and definites predict that the discourse of (10) and (11),

(10) John believes that *a woman*$_x$ broke into his apartment.

---

[176] where $W(X) = \{<g,w> | \exists s: <g,s> \in X \;\&\; s \leq w\}$

(11) John believes that *she*$_x$ is now hiding from the police.

are appropriate if (10) is stated before (11), but not the other way round. The novelty/familiarity problem is solved, the novelty and familiarity condition make sense also for belief contexts.

Asher (1993) notes a problem for the kind of approach I sketched above: two syntactically identical *that*-clauses in the same context, which yield distinct, alphabetical variant formulae under the construction procedure, would have distinct denotations. I believe that this problem is symptomatic for a more serious problem associated with this approach. The problem is to give an intuitive motivation for what it means to say that $x \in \cap\{D(K(a,w))| w \geq s\}$. What does it reflect that *x* is an element of this set? Intuitively it should mean that *the speaker* has used a certain kind of indefinite in an embedded sentence of a belief attribution to characterise the agents belief. However, this is not the kind of answer proponents of constructive update semantics can give. They should say instead that the agent believes a proposition of a particular kind. It seems to me that such an answer shows that we tried to account for intentional identity cases in a conceptually wrong way.

True, if we use a pronoun in an embedded context, this pronoun must have a referent in every possibility in the belief context. But still, the fact that we have only used a limited set of indefinites in belief contexts, and that we have used those indefinites in a certain order, is not a fact about what the agent believes, but a fact about the actual utterance context. The following framework takes this truism more serious.

3.3 The modal subordination account

According to Geurts (1995), the way to solve the novelty/familiarity problem for the one-agent case is to assume that statements involving attitude verbs should be interpreted with respect to special contexts, and also *create* special contexts. For the interpretation of different kinds of embedded sentences, different kinds of contexts might be relevant, and all those contexts should be modelled by a set of world-assignment pairs. Indefinites always introduce 'discourse referents', but the kind of context in which they can do this depends on the verbs and operators used in the sentence. The idea is that formulae representing attitude attributions set up indexed information states and that we can refer back to those information states by anaphoric means. The indexed information states represent the information stated in the embedded clauses of attitude attributions. Just like an original CCT information state, these states are modelled by a set of world-assignment pairs. By having these extra information states around, we can avoid our novelty/familiarity problem. If we are in possibility <g,w> of the main context, and the belief context grows, the assignment-function *g* grows too, because a new *propositional discourse referent* is introduced. The enriched assignment function assigns to the newly introduced propositional discourse referent the context-dependent information expressed by the embedded clause. Accordingly, we can now make use of formulae for representing modal sentences with two extra indices, one of which refers to the information state it anaphorically refers back to, the other to refer to the information state it introduces itself. Thus in 'Bel^{p}_{q} (x, A)', *p* refers back to an earlier information state, and *q* refers to the information state the sentence sets up. I will assume that K(h(x),w) is a set of possible worlds, and W the function from information states to the worlds in this information state, $W(S) = \{w \in W | \exists g: <g,w> \in S\}$. The above formula will be interpreted as follows:

$[[\text{Bel}^{p}_{q}(t, A)]](S) \quad = \quad \{<h,w>: \exists g: <g,w> \in S \ \& \ K(|t|^{g,w},w) \subseteq$

$$W([[A]](g(p))) \ \&\ g[q]h \ \&\ h(q) = [[A]](g(p))\}$$[177]

Desire sentences can be interpreted in a similar way. Assuming that *a wants A* is true in *w* iff in *w*, *a* prefers A above ~A, and that with respect to K *a* prefers A above ~A in *w*, A $<_{a,w}$ ~A iff $\forall w' \in [[A]](K)$: $\forall w'' \in [[\sim A]](K)$: $w' \leq_{a,w} w''$, Geurts' proposed interpretation rule for *want* can be stated in our framework as follows:

$$[[\text{Want}^{p}_{q}(t, A)]](S) = \{<h,w>: \exists g: <g,w> \in S \ \&\ W([[A]](g(p))) <_{||t||^{g,w},w} W([[\sim A]](g(p))) \ \&\ g[q]h \ \&\ h(q) = [[A]](g(p))\}$$

It seems that this approach can handle intentional identity cases in a satisfactory way as long as only one agent is involved. The approach makes it easier to account for anaphoric and presuppositional dependencies across attitude contexts than the more rigid account that Heim (1992) defends. Because all kinds of sentences that give rise to modal contexts are anaphorically dependent on, and introduce themselves new contexts of interpretation, a lot more contexts are available as the contexts of interpretation than Heim allows for. And this is needed, according to Geurts, to account for a sequence like

(12) John wants to write *a book*.
He wants to publish *it* in Holland.

But there are two worries with the modal subordination approach. First, it cannot by itself handle *de re* attributions. Second, it cannot account for intentional identity attributions in a multi-agent case. At first sight it seems that Hob-Nob sentences can be accounted for very straightforwardly in this framework. Sentence (3) is simply translated as Bel^{p}_{q}(Hob, $\exists x$Wx $\wedge$ BBMx) $\wedge$ Bel^{q}_{r}(Nob, KCSx). Unfortunately, this doesn't work: the counterexamples to the laziness account discussed in the introduction of this chapter are also counterexamples to the modal subordination account if used for multi-agent cases. The attribution predicted by the modal subordination account is too strong: Nob does not have to believe that a witch blighted Bob's mare for the sentence to be true, contrary to what is predicted by the modal subordination account.

The laziness account and the modal subordination account have a common problem: they both wrongly predict that the second agent in Hob-Nob attributions must believe everything that is attributed to the first agent. Is it possible to weaken any of those approaches such that this wrong prediction doesn't hold anymore? This will be the issue of the next section.

3.4 Intentional identity by descriptive pronouns only[178]

I have not yet discussed the most obvious way in which at least some cases of intentional identity attributions can be accounted for. Given that we can account for descriptive pronouns that do not have to obey anaphoric island constraints, it seems that the novelty/familiarity problem can easily be solved. We just say that pronouns in embedded clauses of belief attributions must always be descriptive pronouns. If we say that K(a,w) denotes the belief state of *a* in *w*, represented by a set of possible worlds, the interpretation of belief sentences can be given as follows (where the first assignment function assigns properties to variables and the second rigid individual concepts):

$$[[\text{Bel}(t, A)]](S) = \{<g',h,w>|\exists g:<g,h,w> \in S \ \&\ \forall w' \in K(||t||^{g,h,w},w): \exists k$$

[177] Of course, this approach has a foundational problem: it might be that there is a world *w* such that <g,w> ∈ g(p). Fortunately, this problem can be solved; for formal details, see Geurts (1995, pp. 84-85).
[178] Except for the way in which *de re* belief attributions are handled in this section, the proposed solution is basically the same as the one presented in Van Rooy & Zimmermann (1996).

$$<g',k,w'> \in [[A]]\,(\{<g,h,w''>|\ w'' \in K(||t||^{g,h,w},w)\})\}$$

Note that in this way we seem to handle the anaphoric dependence relation for a sentence like

(13) John believes that *the winner of the game* needs to play well,
while Mary believes *he* just must be lucky.

in the intuitively right way. The pronoun *he* is simply used as an abbreviation for *the winner of the game*. Unfortunately, assuming that pronouns in embedded clauses of attitude attributions are always descriptive pronouns, we cannot account for the problems of the laziness account discussed in the introduction of this chapter. Moreover, the approach cannot account for ordinary *de re* attributions, it predicts Ralph to have internally inconsistent beliefs about Ortcutt in Quine's story.[179] But we have seen in chapter 1 that we can solve this latter problem if we make use of counterpart functions.

The interpretation rules for non-belief clauses will be as in the bulk of chapter 2, except that terms are now evaluated also with respect to a counterpart function (where the counterpart functions obey the constraints given at the end of chapter 1):

$$[[Pt_1,\ldots,t_n]]\,(S) = \{<g,h,w,c> \in S |\ <[[t_1]]^{g,h,w,c},\ldots,[[t_n]]^{g,h,w,c}> \in I_W(P)\}$$

$$[[\exists xA]]\,(S) = \{<g'[^x/_{\hat{x}}|\,A\,|^g_h\,],h',w,c>|\ <g,h,w,c> \in S\ \&\ <g',h',w,c> \in \cup_{d\in D}[[A]](S[x:=d])\}$$

Terms are now evaluated as follows (on the assumption that all terms are variables):

$$||t||^{g,h,w,c} = c_W(\$(h(t)(w))),\ \text{if } t \in \text{dom}(h),\ \text{else}$$
$$= \$(g(t)(<w,c>)),\ \text{if } t \in \text{dom}(g) \text{ and } \$(g(t)(<w,c>)) \text{ defined;}$$
$$= \text{undefined otherwise}$$

The abstraction $\hat{x}|\,A\,|^g_h$ used in the interpretation rule for indefinites is that function $f : W \times C \to \wp(D)$ such that:

$$f\,(<w,c>) = \{d \in D(w) \mid [[A]](\{<g,h[^x/_{\mathbf{d}}],w,c>\}) \neq \varnothing\},$$

for any $w \in W$ and $c \in C$.

Now we can redefine the interpretation rules for belief sentences in the following way:

$$[[\text{Bel}(t, A)]]\,(S) = \{<g',h,w,c>|\ \exists g{:}<g,h,w,c> \in S\ \&\ \exists c' \in C_{acq}(||t||^{g,h,w,c},w):$$
$$\forall w' \in K(||t||^{g,h,w,c},w):\ \exists k:\ <g',k,w',c'> \in [[A]]$$
$$(\{<g,h,w'',c'>|\ w'' \in K(||t||^{g,h,w,c},w)\})\}.^{180}$$

We can now account for *de re* belief attributions as in chapter 1, but it is not clear how to account for all intentional identity cases like (3), where the pronoun is not an abbreviation recoverable from the embedded sentence of the belief attribution to the first agent, Hob. In

---

[179] Maybe also the uniqueness condition for *de dicto* pronouns in one-agent cases of intentional identity is too strong.

[180] To get the same results as in chapter 1, we should make a distinction between $[[\ ]]^+$ and $[[\ ]]^-$. This can be done (for the extensional case, see Krahmer 1995), but is not relevant for the analysis of intentional identity attributions. I leave this to the interested reader.

such cases, one might propose, the pronoun is an abbreviation of a description, but a more complicated one than is normally assumed. Let's look at the Hob-Nob sentence again:

(3) Hob believes that a witch blighted Bob's mare, and
Nob believes that she killed Cob's sow.

Let me first say how such a sentence might be represented according to this proposal, and only then discuss why so. The formalisation of the relevant reading of (3) could be

(3e) $\exists eP(e) \wedge \text{Bel}(h, \exists y[\text{Cause}(y,e) \wedge W(y)] \wedge \text{BBM}(y)) \wedge \text{Bel}(n, \text{KCS}(y))$[181]

According to this proposal, intentional identity attributions must be beliefs about something external to both. But the beliefs don't have to be *de re* beliefs about a particular *individual*, the object that was responsible for both of their beliefs can be an *event*, too. The relevant event is singled out by the contextually given predicate *P* in the above formalisation (3e). This is as far as the *de re* account goes. But that's not far enough. The agents must each have a belief object that is somehow related to this event, and moreover, these belief objects must be related to this event in similar ways. How shall we account for that? It is here that descriptive pronouns become relevant. Both agents have a *de re* belief about an event, and they believe that one object was somehow responsible for this event. The above representation of the first conjunct of (3) has the effect that a property is introduced. This property introduced by $\exists y[\text{Cause}(y,e) \wedge W(y)]$ is a function from the way the agent thinks about the denotation of *e* (let's denote that by 'e') to something like 'witch who caused e'. If we are in possibility <g,k,w',c> of the context of interpretation for the embedded clause KCS(y) for Nob, the extension of g(y) at <w',c> will be:

$$\{d \mid \langle d, \|e\|^{g,k,w',c}\rangle \in I_{w'}(\text{Cause}) \text{ and } d \in I_{w'}(W)\}\ .$$

Because $y \notin \text{dom}(k)$ and *y* is not introduced in the main context, the pronoun *she* represented by variable *y* can only be interpreted as a descriptive pronoun. If we presuppose that Nob believes that there is only one witch that caused the relevant event, e, $\exists(g(y)(\langle w',c\rangle))$ will denote the witch that caused the counterpart of this event in w'.

Now, how does this analysis account for the asymmetry problem that cannot be solved by assuming that pronouns are pronouns of laziness? That is, how can we now handle the asymmetry between (5) and (6)? The problem does not arise if it is assumed, as seems natural, that the relevant events with respect to which the concept is introduced for (5) and (6) are, respectively, the event where Smith is shot, and the event where Jones is shot. The asymmetry now simply follows from the fact that while Barsky believes that the one who shot Smith also murdered Jones, Arsky does not believe that the person who shot Jones also murdered Smith.

Although according to the proposal discussed in this section, belief states can simply be modelled by sets of possible worlds, the proposal has its obvious worries. First, we always have to be able to fill in a contextually given predicate P in the *representation* of the belief attribution. It seems that in this way the context-dependence of belief attributions is implemented in a not very satisfying way. The second problem is related with the use of the predicate 'cause' in the embedded clause of the representation. In the scenarios and reports discussed this seems to be what we need, but in general it seems an arbitrary choice; it is not clear that the indefinite should always refer to the causer of the event the sentence is about.

---

[181] In general, it can then be proposed that all sentences of the form *a believes that an S is Q*, where the indefinite has intuitively small scope with respect to the believe predicate, should be represented by something like $\exists eP(e) \wedge \text{Bel}(a, \exists x[\text{Cause}[(x,e) \wedge S(x)] \wedge Q(x))$.

We have seen that the modal subordination approach and the descriptive account are too strong, and in this section we have seen that the weakening of such an approach gives rise to many problems on its own with respect to multi agent intentional identity cases. I belief, however, that even for the one agent case the two approaches are unnatural. Look at the modal subordination account again. It predicts that a sequence of belief attributions of the form Bel^{p}_{q} (a, $\exists$xPx) and Bel^{q}(a, Qx) is already true if for no world compatible with what *a* believes there is no P that is also Q. I guess that the first belief attribution would already be true if for no world compatible with what *a* believes there is no P, but I believe that for an appropriate and truthful use of a pronoun in the second belief attribution, normally something more is required. It seems that to account for anaphoric dependencies across attitude verbs, we should take the notion of *belief object* more seriously than we did until now.

3.5 The domain of quantification is construction dependent

In dynamic semantics, indefinites introduce a specific object to each possibility. It can now be suggested that in case an indefinite is used in an embedded clause of a belief attribution, there still is a specific object that is introduced by the indefinite, but this time it is a *belief object*. How should we model belief objects? It's only natural to assume that belief objects are relevant not only for intentional identity attributions, but also for the analysis of *de re* belief attributions. Hintikka (1969), for instance, has argued that for the analysis of *de re* attributions, we can model 'belief objects' by *individual concepts*, functions from worlds to individuals. Traditionally it was assumed that we should associate with each individual concept a definite description expressible in natural language that denotes at most one individual in every world. Kaplan (1969), however, seems to have some doubts about modelling belief objects by concepts understood this way. Discussing what properties belief objects, or *vivid names* as Kaplan calls them, should have, he argues that

> There are certain features which may contribute strongly to vividness but which I feel we should not accept as absolute requirements. It is certainly too much to require that a vivid name must provide Ralph with a means of recognizing its purported object under all circumstances, for we do not follow the careers of even those we know best that closely.

In the same vein, Perry (1980) wonders what it is for an agent to have a continued belief about an internal 'object of belief'. He argues that it is not a necessary condition for internal continued belief, or for *internal identity*, to have a description in mind that uniquely identifies the object in all circumstances if his beliefs would be true. The arguments of Donellan (1970), Kripke (1972), Putnam (1975), Perry (1977, 1979) and Kaplan (1989) against the description theory of speakers denotation discussed in chapter 1 go in the same direction, of course. Also they have argued that to be able to think or speak about an object, we don't have to have a description in mind such that the object is the unique thing that satisfies this description. All we need is a 'dossier of information' (Evans, 1982) caused by, and associated with the object the belief is about. You might think that thus we should not model belief objects by individual concepts.

But if individual concepts are not suited to stand for belief objects, what then should take their place? Perry (1980) introduced the theoretical notion of a *system of files* in terms of which we can define objects, file cards, that can serve this purpose. Since the introduction of CCT by Kamp and Heim we have got used to such file cards.[182] And, indeed, some proponents of Context Change Theory, like Kamp (1985, 1988, 1990), Asher (1986, 1987), Zeevat (1987) and Spohn (1997), have argued that the theory not only helps to analyse anaphoric relations across attitude attributions, but also suggests an answer to the

[182] See Heim (1982, p. 281) for why we should not take the idea of a system of files too literally. File cards are only defined and theoretical notions.

question what kind of objects speakers refer to when they use indefinites in embedded clauses of attitude attributions. This argument is based on the assumption that not just presupposition states, but also belief states should be modelled as CCT-information states.[183]

According to this view, a belief state, S, should thus be represented by a set of world-assignment pairs. It should represent the truth-conditions not of closed, but of *open* sentences. Let us assume that an information state is represented by a set of world-assignment pairs, and that the domains of all those assignment functions consists of the same set of variables, X, a set of *m* variables. Then, following Dekker (1993), we might say that such an information state contains information about at most *m subjects*. We can define a subject in terms of variables and information states. As mentioned already in chapter 2, we can think of variables as functions from assignments to individuals. We can say that the information associated with variable x is the function [x] from world-assignment pairs to objects such that for all $\langle g,w\rangle \in G \times W$, $[x](\langle g,w\rangle) = g(x)$, if $x \in dom(g)$, and undefined otherwise. If we then limit our set of world-assignment pairs to elements of S, we get what Dekker (1993) called the subject of S stored under x. In terms of the information associated with a variable, we can define the set of subjects of S:

$$\Re(S) = \{[x] \mid \forall \alpha \in S: [x](\alpha) \text{ is defined}\}$$

Note that the set of subjects of S^m is a set of at most *m* total functions in $[S \rightarrow D]$.[184] However, one can also think of it as a set of at most *m partial* functions in $[(G \times W) \rightarrow D]$.

Let's now assume again that a belief state should be modelled by a set of world-assignment pairs, such that the domains of the assignment functions are all the same. Then we can follow Perry and say that the set of belief objects of a belief state is simply the set of subjects associated with this belief state. Then, in turn, we can follow Kaplan (1969) and say that in belief contexts the variables should range over belief objects. Thus, we can propose that the indefinites used in an embedded clause of a belief attribution of a in w can refer only to elements of $\Re(K(a,w))$.

In original CCT, information states are elements of $\wp(G \times W)$. To make sense of the above suggestion in our dynamic framework, we have to change our framework in a number of ways. First, we have to make our possibilities elements of $H \times G \times W$, instead of elements of $G \times W$, where G consists of functions from VAR to D, while H is a subset of $[VAR \rightarrow [[G \times W] \rightarrow D]]$.[185,186]. The reason is that we still want to reflect the intuition that only a few belief objects are introduced in the conversation. In a possibility $\langle g,h,w\rangle$, the first assignment function is always the one of the main context, while $\langle h,w\rangle$ comes from the relevant information state. If a belief object is introduced under variable x, $g(x)$ will be a function that takes a pair like $\langle h,w\rangle$ as argument and gives us an individual, an

---

[183] Somewhat surprisingly, Zeevat (1996, p. 740) claims that discourse referents in DRT/FCS can be interpreted by individual concepts. That is not true, in the way that DRT is normally understood, subjects are more fine grained entities than individual concepts. I will come back to this issue later.

[184] It can be the case that two variables have in every possibility of the information state the same value. An information state that contains information about *m* variables therefore does not have to have *m* different subjects.

[185] This has an unwelcome consequence. Belief states are no longer represented in the same way as presuppositional states. I'm not sure what to say about this.

[186] I simply forget about descriptive pronouns in this section. In the last chapter information states consisted also of possibilities of the form $\langle g,h,w\rangle$. There, one extra assignment function was needed to account for descriptive pronouns. Now one extra assignment function is needed because I assume that belief states are modelled by sets of world assignment pairs.

element of D. The objects introduced are thus of two kinds: normal objects and belief objects. However, if we generalise to the worst case, normal objects will be modelled as special kinds of belief objects, too. The second innovation is that if we want to say that in different contexts indefinites introduce different kinds of objects, we have to make explicit what kind of objects the indefinites introduce in the interpretation rule. The interpretation rule for indefinites will look like this:

$$[[\exists xA]] (S, X) \quad = \quad \cup_{d \in X} [[A]] (S[x:=d], X)$$

where X is the relevant set of 'objects'. An atomic clause will now be interpreted by:

$$[[P(t_1,...,t_n)]] (S, X) \quad = \quad (\{<g,h,w> \in S | <||t_1||_{g,h,w},...,||t_n||_{g,h,w}> \in I_w(P)\}, X),$$

where the terms are evaluated in the following way:

$$||t||_{g,h,w} \quad = \quad g(t)(<h,w>), \text{ if } <h,w> \in dom(g(t)),$$
$$= \quad \text{undefined otherwise.}$$

In extensional contexts, the set X of relevant objects will be the set of rigid and total 'concepts', **D**, defined as $\{s \in [[G \times W] \rightarrow D] | \forall \alpha,\beta \in G \times W: s(\alpha) = s(\beta) \in D\}$. Only for intensional contexts will the set of relevant objects not be such a set of rigid concepts. For the interpretation of a belief clause like Bel(t, A) in context S, we pick for every <g,h,w> in S the belief state of *t* in <g,h,w>, check whether A is true in all possibilities of this belief state, and enrich the assignment function *g* with the belief objects introduced by A:

$$[[Bel(t, A)]] (S, X) \quad = \quad (\{<g',h,w> | \exists g: <g,h,w> \in S \ \& \ \forall \alpha \in K(||t||_{g,h,w},w):$$
$$<g',\alpha> \in [[A]] (\{<g,\alpha'> | \alpha' \in K(||t||_{g,h,w},w)\}, \Re(K(||t||_{g,h,w},w)))\}, X)$$

It is in this interpretation rule that the belief objects become relevant. You might think of the interpretation rule for belief sentences as having an effect such that a formula like 'Bel(a, $\exists$xA)' is really interpreted as if it were of the form '$\exists$xBel(a, A)'. But there is an important difference between the two fomulae: in the latter case the object introduced is an ordinary existing individual, while in the former case the object introduced is a belief object of *a*.

Note that in this way it becomes almost impossible for variables introduced in a belief context to be referred back to in the main context, which is clearly a pleasant consequence of this approach. And another one can be seen by looking back at constructive update semantics: there, too, belief states were crucially represented by sets of world-assignment pairs. But in that approach, the variable used *by the attributer* by which an indefinite is introduced must be the same as the variable by which *the believer* has beliefs about the object. That would be a great coincidence, and it is better not to trust on so much luck. In the approach we take now, it's irrelevant under what variable the believer stored his subjects.[187] The variable used by the attributer doesn't have to be the same.

Let us now see how the above proposal accounts for the specificity problem and the problem of intentional identity. It is easy to see that belief attributions with indefinites occurring in embedded clauses are not necessarily about specific objects. If a formula like Bel(a, $\exists$xPx) is accepted in possibility <g,h,w>, it might well be that there are possibilities α and β in K(a,w) such that $\alpha(x) \neq \beta(x)$. Subjects need not be rigid functions. It goes

[187] As a result, the information contained in a belief state are closed under permutation of variables. Although the information states are structured, the approach is not representational.

without saying that for the one agent case, intentional identity is also unproblematic in case both the indefinite and the pronoun occur in a belief clause. But what about intentional identity cases where more agents are involved?

3.6 Syntactic counterparts and common grounds

A counterpart relation is sometimes thought to be of help here.[188] It is normally assumed that two belief objects of different agents can be counterparts of each other for two reasons. Either each agent has a belief object that is the other's counterpart by *communication*; or these belief objects are counterparts by *experience*, because they are derived by whatever means from the same source. One might assume that two belief objects are counterparts of each other - *syntactic counterparts* - if they are stored under the same variable. Two belief objects of different agents can be counterparts by communication, because the two agents have communicated with each other about a certain object, and for that reason parts of what they believe have a common ground. We can define the common ground of two CCT information states S and S' in the manner proposed by Dekker (1993):

$$S \vee S' \quad = \quad \{\langle g,w\rangle \mid g \in D^{D(S)\cap D(S')} \,\&\, \langle g,w\rangle \ll S \text{ or } \langle g,w\rangle \ll S'\},$$

where $\langle g,w\rangle \ll S$ iff $\exists \langle h,w'\rangle \in S$: $g \subseteq h$ & $w = w'$.

Simplistically, we might assume that if two agents have talked with each other about a certain object in indexed English, they have both represented this object by the same variable. We can now suggest that in case of Hob and Nob, the indefinite *a witch* used in the embedded clause of the belief attribution for Hob does not introduce merely a belief object of Hob, this object is also a *shared* belief object of Hob and Nob. For the intentional identity attribution to make sense, the subject introduced by *a witch* should be not only an element of $\Re(K(Hob,w))$, but also an element of $\Re(K(Hob,w) \vee K(Nob,w))$. If the latter condition is not satisfied, the belief attribution to Nob would not be appropriate.

Let's consider again the Edelberg sentences that gave rise to the asymmetry problem.

(5) Arsky believes that someone murdered Smith, and
Barsky believes he murdered Jones.
(6) Barsky believes that someone murdered Jones, and
Arsky believes he murdered Smith.

What we need to explain is that in cases 3 and 5 given in the quotations of Edelberg (1995), (5) is true and (6) is false on their most natural readings. Unfortunately, this cannot be explained with the machinery we have available. Let's assume that x and y are the variables under which Arsky and Barsky have stored their belief subjects of, respectively, the murderer of Smith and the murderer of Jones. Because Arsky has a two-murderer theory, in all possibilities $\langle g,w\rangle$ representing Arsky's belief state, $[x](\langle g,w\rangle) \neq [y](\langle g,w\rangle)$. Barsky, instead, has a one-murderer theory. It follows that in all possibilities $\langle h,v\rangle$ representing his belief state, $[x](\langle h,v\rangle) = [y](\langle h,v\rangle)$. The reason why (5) is true is immediately clear. The clause *someone murdered Smith* picks up Arsky's belief object stored under x; and in all possibilities representing the belief state of Barsky, x and y denote the same individual. The problem is that the falsity of (6) cannot be explained. The clause *someone murdered Jones* can pick up Barsky's subject stored under y, but also the subject stored under x. Because of the latter possibility it is falsely predicted that the second conjunct of (6) is also true.

At this point Edelberg (1992, 1995) intervenes. He argues that for Barsky also the belief objects associated with *the murderer of Smith* and *the murderer of Jones* should be distinct,

[188] Kamp (1985, 1988, 1990), Asher (1987), Zeevat (1987), Edelberg (1992, 1996).

even though Barsky believes that they stand for the same individual. We can formalise this in our framework by saying that belief objects are not modelled by subjects, but by the *variables* that give rise to such subjects.[189] Thus, a formula of the form $\exists xA$ inside the scope of a belief operator introduces a variable, say z; and the interpretation of x in $<g,h,w>$ will be $h(g(x))$, if $g(x)$ is a variable. But this by itself won't help, because the indefinite in (6) can still denote both x and y. Indeed, Edelberg argues that to account for the asymmetry problem, we have to assume that indefinites can be used not only quantificationally, but also *specifically*. In the case at hand, the truth conditions of (5) and (6) on their most natural readings differ from each other, because (in terms of the present framework) in (5) the belief object stored under x is introduced by the indefinite, and in (6) the belief object stored under y. This might well be the right conclusion (see § 3.9), but in this chapter my main concern will be to see how well intentional identity attributions can be handled in the DRT/FCS framework.

We have seen that in this framework it doesn't help much to model belief objects by subjects. Modelling belief objects by the variables themselves is not only suspiciously syntactic, but doesn't help anyway from a DRT/FCS point of view. What should we do then?

Perhaps we should look not at the subjects of the belief states of the agents themselves, but rather at the subjects in the *common ground* of the belief states of Arsky and Barsky.[190] Although Barsky has a one-murderer theory, Arsky does not, so also there is no assumption of only one murderer in the common ground. In general, then, according to this new proposal, belief attributions are interpreted as follows:

$$[[\text{Bel}(t, A)]](S, X) = (\{<g',h,w> \mid \exists g:<g,h,w> \in S \;\&\; \forall \alpha \in K(|t|_{g,h,w},w): <g',\alpha> >> [[A]](\{<g,\alpha'> \mid \alpha' \in CG(w)\}, \Re(CG(w)))\}, X)$$

where $CG(w)$ is the relevant common ground, and $<g,\alpha> >> S$ iff for all $<h,\beta> \in S$: $g \geq h$ and $\alpha = \beta$. In the case of Arsky and Barsky, the relevant common ground, $CG(w)$, will of course be the common ground of Arsky and Barsky in w: $K(\text{Arsky},w) \vee K(\text{Barsky},w)$.

Unfortunately, this proposal by itself won't do. As can easily be seen, the asymmetry between (5) and (6) is still not accounted for. The reason is that if [x] restricted to CG represents the one who shot Smith, this subject can be picked up as the referent of the indefinite *someone* in (6) and would make both conjuncts of (6) true because Barsky has a one-murderer theory. Paul Dekker (personal communication) has proposed that if we interpret embedded sentences of belief attributions with respect to the common ground of more agents, it is natural to distinguish in the embedded sentence the part that is already common ground, the *background*, from the part that is new, the part that stands in *focus*[191]. Perhaps in this way the asymmetry can be accounted for. To implement this proposal, I will assume that the background is presupposed material, which will be indicated by Beaver's (1993) presupposition operator, ∂. The old information given by a sentence, the background, is now separated formally from the new information, the focused part: sentences are now formalised as conjunctions of the form $\partial(A_1) \wedge A_2$, where ∂A is interpreted as follows:

---

[189] To use the variables themselves to model belief objects was suggested to me by Paul Dekker (personal communication).

[190] I think that something like this is what Kamp (1985, 1990) had in mind.

[191] See Von Stechow (1990) for a more formal way of working out this kind of approach with respect to focus-background structure, in the framework of structured propositions.

$[[\partial A]](S, X) = [[A]](S, X)$, if $S \copyright \text{first}([[A]](S, X))$[192],
$= \varnothing$ otherwise

which means that A may introduce new discourse markers, but does not contain new 'worldly' information with respect to S.

To let the background do some work, we can represent our problematic sentences (5) and (6) as follows:

(5') $\text{Bel}(a, \partial(\exists x Px) \wedge M(x,s)) \wedge \text{Bel}(b, M(x,j))$

(6') $\text{Bel}(b, \partial(\exists x Px) \wedge M(x,j)) \wedge \text{Bel}(a, M(x,s))$,

where P is a contextually given predicate. Obviously, P cannot express a trivial property; for that would leave the asymmetry unexplained. In the case at hand, something like *shot Smith*, for (5'), or *shot Jones*, for (6'), is called for. In this way, $\exists x Px$ does not give rise to new 'worldly' information with respect to the common ground of Arsky and Barsky, and the asymmetry can be explained. In (5'), the subject corresponding to the description *the one who shot Smith* is introduced; and it is claimed that Arsky believes that he murdered Smith, and that Barsky believes that he also murdered Jones. According to Edelberg's scenario, (5') is predicted to be true. For (6'), the subject corresponding to the description *the one who shot Jones* is introduced; and it is claimed that Barsky believes that he murdered Jones, and that Arsky believes that he murdered Smith. According to Edelberg's scenario, Arsky doesn't believe that the one who shot Jones is also the one who murdered Smith, thus (6') is predicted to be false, as desired.

I believe that this proposal is natural insofar as we do interpret intentional identity attributions with respect to the common ground of the relevant agents. But this leaves this proposal with a problem similar to the problem of the proposal discussed in § 4: we always have to be able to fill in a contextually given predicate P. The analysis is problematic not because what is expressed by the belief attribution is made context-dependent, but because this context dependence is implemented by making the logical *representation* of the attribution depend on context. Another problematic feature of the proposals discussed in this section, is that it is assumed that two belief objects can be counterparts of each other, only if they are stored under the same variable. In the next section I will try to account for the context dependence of intentional identity attributions without making the representation of the sentence dependent on context, and without assuming that 'counterparthood' is a syntactic matter. In this way we won't have to refer explicitly to the relevant common ground.

3.7 Counterparts and externalism

How should we account for intentional identity if we don't refer to the relevant common ground? What we should do in this case becomes clear when we look at another problem for the accounts sketched in the last sections. The above accounts cannot analyse *de re* belief attributions in an appropriate way. The reason is that when in the main context only real individuals are introduced, it is not clear how to account for Quine's *double vision* problem of *de re* belief attributions. The most natural way to go would seem to be to follow Kaplan (1969) and assume a representation relation between subjects, real world objects, and agents, such that the subject stands in this relation to the object for the agent, if the agent is acquainted with this object under this subject; and the subject is defined for all elements of the belief state of the agent. But now Edelberg (1992, 1995) complains that such an account is in need of two kinds of counterpart relations: one kind of counterpart relation between subjects of different information states, and another kind of relation

[192] Where $S \copyright S'$ iff for every $\alpha \in S$ there is an $\alpha' \in S'$: $\alpha \leq \alpha'$, and where first(A.B) = A.

between subjects, real world objects, and agents.[193] Edelberg suggests that we had better forget about the last kind of relation and try to account for *de re* belief attributions also in terms of the former kind of counterpart relation. I agree that if we can account for intentional identity attributions and *de re* belief attributions with the help of only one kind of counterpart relation, then this should be preferred to an analysis that needs counterparts and representation relations. Edelberg shows that we can account for both kinds of belief attributions in a perspectival semantics by using only counterpart relations between belief objects, and argues that a perspectival semantics is to be preferred to a realist analysis. With this latter claim I don't agree, however. In this section I will argue that the realist dimension is crucial for a proper account of intentional identity attributions. Moreover, I will show that intentional identity attributions and *de re* belief attributions can be accounted for in a similar way in a realist semantical account. We need only one kind of counterpart relation to handle both kinds of attributions. However, this kind of counterpart relation will be crucially different from the ones discussed in the last section.

The account until now was simplistic for an obvious reason. It was built on the assumption that a belief object of one agent is a counterpart to a belief object of another agent if and only if the two agents have stored their belief objects under the same variable. But, of course, there is no reason at all to make this assumption. To be less simplistic, it seems better to make use of a primitive counterpart relation between belief objects in the model that is independent of how these belief objects are stored. It can then be assumed that in the embedded clause of a belief attribution in which a pronoun occurs, the interpretation of the variable by which this pronoun is represented does not necessarily refer to the same belief object as the actual referent of the variable, but has to refer to a belief object that is a counterpart of the actual referent of this variable.

To account for the *double vision* problem of *de re* belief attributions discussed by Quine (1956), I have proposed in chapter 1 that the problem in question can be solved by assuming that a single object in the real world might have two different *counterparts* in worlds that characterise a belief state, because there might be different ways of picking out counterpart relations. We can say that a counterpart relation is really a function, that given an individual d and a world w, will give us an individual d' in w. A *de re* modal statement is now absolutely true (false), if it is true (false) with respect to all conversationally relevant ways of picking out counterparts - that is, with respect to all conversationally relevant counterpart functions. In other words, for a semantic account of *de re* modal statements, we rely on Van Fraassen's (1966) notion of *supervaluation*. What is attractive about this approach is that it can account for the intuition that what is expressed by a modal statement is very context dependent, but still leaves the *semantics* of such statements relatively simple. The idea is that we can rely on this simple semantics if we analyse modal statements out of context. One might suggest that this kind of approach is not only successful for the analysis of *de re* modal statements, but might also be useful for the analysis of intentional identity attributions. We have to change it in one way, though. Obviously, to account for intentional identity, counterpart functions can no longer be functions from individuals and worlds to individuals, but must be functions from belief objects and information states to belief objects.

Although there is nothing wrong with assuming the existence of primitive counterpart relations, we would still like to know how to explain these relations. We have assumed above that there are two ways in which two belief objects of different agents might be counterparts of each other: either their respective belief objects are derived by being in sensory contact with the same object, in which case they are *counterparts by experience*; or their belief objects have become counterparts through *communicative* links. Indeed, with these two kinds of counterpart relations we can explain a lot of true and appropriate intentional identity cases where more agents are involved. Unfortunately, not all such cases can be explained in this way. Consider the following example from Edelberg (1992):

---

[193] Asher (1987) has indeed made a distinction between *external anchors* and *quasi-external anchors*.

Edelberg's Astronomers

Two teams of astronomers have independently been investigating the peculiar motion of superclusters of galaxies - that is, the motion of *clusters* of clusters of galaxies, over and above that due to the Hubble expansion of the universe. Neither team knows about the work of the other, but both independently and correctly ascertain the peculiar motions of the Hydra-Centaurus supercluster, of the Local Supercluster, and of our own Local Group. Both teams attempt to explain the vectors of the peculiar motions in the same way: by postulating an "overdensity" of galaxies at roughly twice the distance between the Hydra-Centaurus supercluster and our own galaxy. The idea is that an enormous collection of galaxies, "a distant concentration of mass that appears to be larger than any proposed by existing cosmologies," lies beyond the Hydra-Centaurus supercluster, drawing it, as well as our own Local Supercluster of galaxies, towards it. The American team calls the structure "The Great Attractor", the Soviet team calls it "The Overdensity" (in Russian). Due only to certain differences in instrumentation and atmospheric conditions at the times and locations of observations, the two teams conjecture the structure to be at "slightly" different distances. The Americans say it is twice the distance of the Hydra-Centaurus supercluster; the Soviets say it is at 2.1 times the distance. In reality, let us suppose, the Great Attractor does not exist at all: the peculiar motions of the various superclusters are each caused by independent factors.

In this situation, the following intentional identity statement would be true:

(14) The American team believes that *an immense overdensity* of galaxies certain distance causes a peculiar motion of superclusters, but the Soviet team thinks *it* is slightly further away.

So, in this case we have intentional identity, although the belief objects can be neither counterparts by experience derived from an existing object, nor counterparts by communication. Note also that interpreting the pronoun as a descriptive pronoun won't help.

It thus seems that our account of the counterpart relation does not always work. We may therefore try to follow Edelberg (1992), who proposes to explain counterparts in terms of *rough similarity of explanatory role*.[194] There certainly are cases where this is all that is required. For the majority of intentional identity cases, however, 'having a same concept', or 'having a subject that internally plays a similar role' is not enough. To see this, consider a Twin Earth variant of the above Astronomers story, according to which the two teams live in two distinct regions of the universe that happen to be qualitatively identical. As Zimmermann (forthcoming) observes, in that case (14) seems to be false, although just as in Edelberg's own story the relevant belief objects play a roughly similar explanatory role. With Zimmermann, I conclude that rough similarity of explanatory role cannot be a sufficient condition for intentional identity. Discussing the epistemic role of discourse referents, Zimmermann (1995) proposes that subjects represent *sources* of information.[195] It seems only natural to propose that two subjects are counterparts of each other if they represent the same source; or at least that it is a minimal condition for two subjects, or belief objects, to be counterparts of each other that they have their *source* in the same event. Note that this constraint on counterpart relations can be formulated only in a realist semantic account. Thus, if it holds that the belief objects (if any) relevant for intentional identity attributions must have their source in the same event for the intentional identity attribution to be true and appropriate, then this rules out, I think, a perspectival analysis of such ascriptions. Let us now try to account for intentional identity cases in terms of a realist semantic theory by means of counterpart relations that obey the above source condition.

---

[194] Compare this with Lewis's (1968) account of counterpart relations by means of qualitative similarity.

[195] In a similar vein, Stalnaker (1987b, 1988) argues that counterpart relations between 'individuals' living in different worlds should not be explained in terms of qualitative similarity and difference, but instead by means of the *causal relations* between real world objects and the representations of those objects by the relevant believer.

Given a set of worlds, W, a designated element of W, w_0, representing the actual world, a non-empty set D of objects, and a set VAR of variables, there is a set G of partial *assignments* associated with D. The set H is a set of partial assignments associated with VAR and $[(G \times W) \rightarrow D]$. We assume that K(a,w) is the CCT information state of *a* in *w*.

If S is a CCT information state, we can define, following Dekker (1993), the information of S associated with variable *x* as:

$[x]_S :=$ the function $f \in [S \rightarrow D]$ such that $\forall \alpha \in S: f(\alpha) = \alpha(x)$

In contrast to the subjects used in earlier sections, subjects as defined here *live in* information states. But to account for intentional identity, there might exist counterpart relations between subjects living in different information states. In distinction with the subjects defined in section 4 and used in section 5, the subjects as defined above live in information states. The set of subjects of an information state is defined much as before: $\Re(S) := \{[x]_S | x \in dom(S)\}$. In chapter 1, a counterpart function was a function from objects and worlds to objects. Now a counterpart function is a function from subjects and information states to subjects. I will introduce the distinguished subject, †, and will assume that for all $<g,w> \in [G \times W]$, $\dagger(<g,w>) = *$. Thus CP(w), the set of admissible counterpart functions between subjects in *w*, is a subset of $[[[(G \times W) \rightarrow D] \times \wp(G \times W)] \rightarrow [(G \times W) \rightarrow D]]$. If s is a subject and Z an information state, cp(s, Z) will denote the subject that is the counterpart of s under cp in information state Z. For every w, the set CP(w) obeys the following constraints:

For all $cp \in CP(w)$, $s \in [(G \times W) \rightarrow D] \cup \{\dagger\}$ and $Z \in \wp(G \times W)$:

(a) if $s \in \Re(Z)$, then $cp(s, Z) = s$

(b) $cp(s, Z) \in \Re(Z) \cup \{\dagger\}$[196]

(c) $cp(\dagger, Z) = \dagger$[197]

Two subjects can be counterparts of each other only if they have the same source. I implement this by requiring that the counterpart functions have to obey the following constraint for all worlds w:

For all $cp \in CP(w)$, $s,s' \in [(G \times W) \rightarrow D]$ and $Z \in \wp(G \times W)$:

if $cp(s, Z) = s'$, then there is an event $e \in D(w)$ such that s and s' have common source *e*.[198]

Possibilities will now be elements of $(H \times (G \times W) \times CP \times \wp(G \times W))$, and are ordered by $\leq$: if β and β' are elements of $(G \times W)$, then $<g, \beta, cp, Z> \leq <h, \beta', cp', Z'>$ iff $g \subseteq h$

---

[196] Note that if $Z \neq G \times W$, no element of **D** is an element of $\Re(Z)$.

[197] Note the similarity between the constraints on the counterpart functions between subjects defined here, and constraints on the counterpart functions between objects in chapter 1.

[198] This doesn't mean that in all true cases of intentional identity we are talking about two belief objects that have the same source. Sometimes we don't seem to talk about belief objects at all, but account for intentional identity by descriptive pronouns. This seems to be the case, for instance, in *John believes that the winner of the game needs to play well, while Mary believes he must only be lucky*. In other cases belief objects do seem to play a role, although the belief objects are not grounded in the same event. Consider the two children Pierre and Ralph. Pierre is French, Ralph is English, and they do not know each other. They both believe in the existence of a character we call *Santa Claus*, but Pierre only 'knows' him under the name *Père Noël*. In that case, intuitively, the following intentional identity statement is true, *Pierre believes that Père Noël comes to France only at Christmas, and Ralph believes that he brings presents*, although the beliefs of Pierre and Ralph are not based on a single event.

& $\beta = \beta'$ & cp = cp' & Z = Z'. This ordering relation caries over to information (presupposition) states S and S': $S \leq S'$ iff for every $\alpha' \in S'$ there is a $\alpha \in S$: $\alpha \leq \alpha'$. Information state-domain pairs also stand in this ordering relation: $(S, X) \leq (S', Y)$ iff X = Y and $S \leq S'$.

For the interpretation of negation I introduce the notion of *subtraction*. Subtracting state S' from state S, S - S', will leave us with those elements of S that have no extension in S': S - S' = $\{\alpha \in S \mid \neg\exists\alpha' \in S': \alpha \leq \alpha'\}$. In the same way, subtracting (S',Y) from (S,X), (S,X) - (S',Y), is (S - S', X).

As before, I will assume that in the main context we are always talking about real world objects, elements of **D**.[199] But because every term is evaluated with respect to a counterpart function, I assume that the default information state is DIS, for which it holds that for every cp ∈ CP: (i) for all **d** ∈ **D**: cp(**d**, DIS) = **d**, and (ii) for all s ∈ [[(G × W) → D], if s ∉ **D** then cp(s, DIS) = †.

If E is [(G × W) → D], and X is an arbitrary subset of E, I can now give a recursive definition of the context change potential $[[A]] [\subseteq <\wp(F \times (G \times W) \times CP \times \wp(G \times W)) \times \wp(E)> \times <\wp(F \times (G \times W) \times CP \times \wp(G \times W)) \times \wp(E)>]$ of formulae A:[200]

(1a) $[[P(t_1,\ldots,t_n)]] (S,X) = (\{<g, h, w, cp, Z> \in S \mid <\|t_1\|_{g,h,w,cp,Z},\ldots,\|t_n\|_{g,h,w,cp,Z}> \in I_W(P)\}, X)$

(1b) $[[t_1 = t_2]](S,X) = (\{<g, h, w, cp, Z> \in S \mid \|t_1\|_{g,h,w,cp,Z} = \|t_2\|_{g,h,w,cp,Z}\}, X)$

The term-evaluation used in (1a) and (1b) is defined with respect to a counterpart function as follows:

$\|t\|_{g,h,w,cp,Z}$ = cp(g(t), Z)(<h,w>), if $t \in dom(g)$,
= undefined otherwise.

On the basis of this, we can define the interpretation rules for complex formulae:

(2) $[[\neg A]] (S, X) = ((S, X) - [[A]] (S, \mathbf{D}))$[201]

(3) $[[A \wedge B]] (S, X) = [[B]] ([[A]] (S, X))$

(4) $[[\exists xA]] (S, X) = \cup_{s \in X}[[A]] (S[x:= s], X)$[202]

(5) $[[Bel(t, A)]] (S, X) = (\{<g', h, w, cp, Z> \mid \exists g:<g, h, w, cp, Z> \in S \ \& \ \forall \alpha \in K(\|t\|_{g,h,w,cp,Z,w}): <g', \alpha, cp, K(\|t\|_{g,h,w,cp,Z,w})> \in [[A]]$

---

[199] In the terminology of Edelberg (1992, 1995), the framework is a realist one, and not perspectualistic, as the one of Edelberg.

[200] Obvious definedness conditions have been omitted for ease of exposition.

[201] So, indefinites used under the scope of a negation do not 'introduce' belief objects of the relevant agent. With counterparts, though, there is still a problem in cases where negation takes scope over pronouns. This problem can be solved by also letting negation change the information state of the possibilities from any Z into the default information state, DIS, and by giving a slightly different definition of subtraction. I ignore this issue here, however.

[202] S[x:=s] is defined analogously to the definition of S[x:= d] in chapter 2: $S[x:= s] := \{<g', h, w, cp> \mid \exists g:<g, h, w, cp> \in S \ \& \ g[x]g' \ \& \ g'(x) = s\}$.

$$(\{<g, \alpha', cp, K(|lt|g,h,w,cp,Z,w)>|\alpha' \in K(|lt|g,h,w,cp,Z,w)\},$$
$$\Re(K(|lt|g,h,w,cp,Z,w)))\}, X)$$

Now we can say that formula A is *acceptable* in S, $S\models A$, if S is a *substate* of [[A]] (S), in the sense that for every $\alpha \in S$ there is an α' in [[A]] (S) such that $\alpha \leq \alpha'$. As in chapter 1, I want to use a *supervaluation* account of truth. A sentence is only true, if it is true for all admissible ways of picking out counterparts. Thus, A is *true* with respect to g, h, w, and Z, $<g, h, w, Z>\models A$, iff for all $cp \in CP(w)$: $\{<g, h, w, cp, Z>\}\models A$. Similarly, A is *false* with respect to g, h, w, and Z, $<g, h, w, Z>=| A$, iff for all $cp \in CP(w)$: $\{<g, h, w, cp, Z>\}\not\models A$.[203]

Just like in the case of *de re* belief attributions we can account for the fact that in the story told by Quine the belief attribution *Ralph believes that Ortcutt is a spy* is strange by means of supervaluation, now I want to argue that by means of supervaluation we can account for Edelberg's asymmetry problem.

Why is supervaluation helpful in accounting for the asymmetry problem? Look at the case of Arsky and Barsky again. Because Arsky has a two-murderer (or -shooter) theory, and Barsky only a one-murderer (or -shooter) theory, we can say that Arsky has two relevant belief objects, [S] and [J]; while Barsky has only one relevant belief object, [SJ]. If we account for intentional identity by means of counterpart functions, there are only two relevantly different counterpart functions, cp and cp'. Let K(a) be Arsky's belief state, and K(b) be Barsky's belief state. Intuitively it is the case that Barsky's one belief object, [SJ], is in K(b) a counterpart of Arsky's [S] and [J], for any counterpart function. That is, cp([S], K(b)) = [SJ], cp([J], K(b)) = [SJ], cp'([S], K(b)) = [SJ] and cp'([J], K(b)) = [SJ]. This is the reason, in the present framework, that (5) is acceptable and true in the situation sketched by Edelberg:

(5) Arsky believes that someone murdered Smith, and
Barsky believes that he murdered Jones.

On the other hand, it is the case neither that for all counterpart functions, Arsky's [S] is in K(a) a counterpart of Barsky's [SJ]; nor that for all counterpart functions, Arsky's [J] is in K(a) a counterpart of Barsky's [SJ]. That is, although one counterpart function, say cp, will relate [SJ] to [S] in K(a): cp([SJ], K(a)) = [S], the other counterpart function will relate [SJ] to [J] in K(a): cp'([SJ], K(a)) = [J]. But in that case, (6)

(6) Barsky believes that someone murdered Jones, and
Arsky believes that he murdered Smith.

will not be true and acceptable anymore for all ways of picking out counterparts in the situation sketched by Edelberg.

Let <g, h, w, Z> be any possibility compatible with everything *we* know about what Arsky and Barsky believe about the cases of Smith and Jones - accordingly, that Arsky has a two-murderer theory, and Barsky a one-murderer theory. In that case, (5") will be true in <g, h, w, Z>, but (6") will not:

---

[203] Of course we can also implement supervaluation in the dynamic meaning. In that case, [[ ]] will no longer be the only interpretation rule for sentences. For every natural language sentence, and not just any sentential clause, the most important interpretation rule is given by [.]; and for this interpretation rule, and only for this interpretation rule, *supervaluation* will be important. Now we can say that <g, h, w, Z> is an element of the first set of [A](S, X) iff for every $cp \in CP(w)$:<g, h, w, cp, Z> is an element of the first set of [[A]] (S, {<k, l, w', cp', Z">| $\exists$<k, l, w', Z'> $\in$ S & cp' = cp & Z' = Z"}, X).

(5") $Bel(a, \exists xMSx) \wedge Bel(b, MJx)$

(6") $Bel(b, \exists xMJx) \wedge Bel(a, MSx)$

On the above *semantic* account, the truth value of a sentence depends via supervaluation on all possible ways of picking out counterparts. However, *pragmatically speaking* it might be that a certain sentence is counted as true because it is true with respect to all the *conversationally relevant* counterpart functions, although it is not true for all possible ways of picking out counterparts. Thus, in the pragmatics of belief attributions we look only at all the conversationally relevant counterpart functions. We need to make what is expressed by a belief attribution context-dependent in this way, because we also have to account for the *true* uses that (6), (15) and (16) have in Edelberg's example:

(15) Barsky thinks that someone murdered Smith,
and Arsky thinks that *he didn't* murder Jones.

(16) Barsky thinks that someone murdered Smith,
and Arsky thinks that he is still in Chicago.

Of course, the proposed account predicts that (6), (15) and (16), though they can be true, can be counted as true only after some extra pragmatic reasoning. The context has to be accommodated in such a way that only some specific counterpart functions are relevant for the interpretation.

Edelberg (1992, 1995) has claimed that a realist semantics cannot account for *de re* attributions in terms of the same counterpart relations used for the analysis of intentional identity attributions. On the account presented here, though, we can. The reason is that objects, elements of D, are treated as special kinds of subjects. If we represent a *de re* belief attribution like *Barsky believes that Arsky is a detective* by the formula $\exists x[x = a \wedge Bel(b, Dx)]$, the sentence will be true with respect to counterpart function *cp* if and only if Barsky believes that the subject that is the counterpart of the rigid function from possibilities to Arsky under *cp* is a detective.

3.8 Belief objects as individual concepts

The analysis of the last section was still based on the assumption that we should model belief states by sets of world-assignment pairs, and belief objects by subjects. The background idea was that we would not be able to single out belief objects in a fine grained enough way, if we would represent a belief state by a set of possible worlds. In the beginning of §3.4 I suggested that the arguments of Kaplan and Perry ruled out the idea that individual concepts can play the role of belief objects. They seemed to be too coarse grained entities. But I don't think that the argument given there was convincing. What Perry argued, for instance, was that for internal identity is it not needed that the agent has a specific *description* in mind. It is clear that we should not assume that a belief object should be thought of as, or can be associated with, a descriptions that determines a unique referent in all possibilities that help to represent the agents belief state. Subjects are not thought of in that way, but neither do we have to think of individual concepts in this way. In fact, once we assume that individuals in different possible worlds can be 'identified' by means of a primitive counterpart relation, as we did in chapter 1, we can construct individual concepts that are not associated with any descriptive material, without relying on any notion of qualitative similarity.

If we want to account for the notion of *belief object* in terms of individual concepts, we can represent a belief state by a pair like <K, BO>, where K is a set of possible worlds, and BO is a set of functions in $[K \rightarrow D]$.

To account for the asymmetry problem in the same kind of way as we did in §3.6, we should now think of counterpart functions as functions from concepts and sets of worlds to concepts.[204]

Atomic formulae are now interpreted as in the above given subjectival semantics, except that the possibilities are not five, but quadruples, and that the last element of this tuple is now a set of possible worlds:

$$[[P(t_1,\ldots,t_n)]](S,X) = (\{<g, w, cp, Z> \in S \mid <||t_1||_{g,w,cp,Z},\ldots,||t_n||_{g,w,cp,Z}> \in I_w(P)\}, X)$$

In the above interpretation rule, the terms are evaluated as follows:

$$||t||_{g,w,cp,Z} = cp(g(t), Z)(w)$$

In the clause for the interpretation of belief sentences, only a limited set of individual concepts can be introduced, and the terms of the embedded clause are evaluated with respect to the information state of the agent:

$$[[Bel(t, A)]](S, X) = (\{<g', w, cp, Z> \mid \exists g{:}<g, w, cp, Z> \in S \;\&\; \forall w' \in K(||t||_{g,w,cp,Z},w){:}\ <g', w', cp, K(||t||_{g,w,cp,w})> \in [[A]](\{<k, w'', cp', K(||t||_{g,w,cp,Z},w)> \mid w'' \in K(||t||_{g,w,cp,Z},w)\}, BO(||t||_{g,w,cp,Z},w))\}, X)$$

Just as before, only in case indefinites are used in the scope of belief predicates we might introduce non-rigid concepts. For extensional contexts the set of relevant objects X is always the set of rigid individual concepts, and the relevant information state Z will be the analogue of DIS of §3.6. It should be obvious that Edelberg's asymmetry problem will be tackled in the same way as we did in the last section: absolute truth and falsity is defined in terms of truth and falsity with respect to all counterpart functions.

It might seem that working with individual concepts instead of subjects has no real advantage. It's true that we don't have to represent belief states by sets of world-assignment pairs, but, so it seems, only at the cost of representing a belief state by a tuple like <K, BO>, where K is a set of worlds and BO a set of individual concepts representing the belief objects. But given the counterpart theory of chapter 1 introduced to analyse *de re* belief attributions, I want to argue now that this is not really the case. For each agent *a* and world *w*, we can define BO(a,w) in terms of K(a,w) and the set of counterpart functions C(a,w), which is a superset of $C_{acq}(a,w)$ as given in chapter 1. If K_0 and C_0 are the abbreviations for respectively K(a,w) and C(a,w), and if we abbreviate '$\lambda w.c_w(d)$" by "c[d]" the definition goes as follows:

$$BO(a,w) := \{f \in [K_0 \rightarrow D] \mid \exists c \in C_0: \exists d \in \bigcup\{D(w') \mid w' \in K_0\}:$$

---

[204] For any w, the set CP(w) of counterpart functions in w should now obey the following constraints:
For all cp $\in$ CP(w), ic $\in [W \rightarrow D \cup \{*\}]$ and K $\in \wp(W)$:
(a) if ic$\restriction$K $\in [K \rightarrow D]$, then cp(ic, K) = ic$\restriction$K;
(b) cp(ic, K) $\in [K \rightarrow (D \cup \{*\})]$;
(c) if ic = W × {*}, then cp(ic, K) = ic.
Also, two concepts can only be counterparts of each other if they have the same source. I implement this by demanding that the counterpart functions have to obey the following constraints for all w:
For all cp $\in$ CP(w), ic,ic' $\in [W \rightarrow D]$ and K $\in \wp(W)$: if cp(ic, K) = ic', then there is an event e $\in$ D(w) such that ic and ic' have common source e.

$$\forall w,w' \in K_0: c_w(c_{w'}(d)) = c_w(d) \;\&\; f = c[d] \restriction K_0]^{205}$$

The above definition guarantees that for each concept f of BO(a,w), the values of f in the worlds of K(a,w) are all counterparts of an individual in a world in K(a,w). Note that the concepts in BO(a,w) are only defined for worlds in K(a,w). This is needed to guarantee that a sequence like

(17) John believes that *a woman* broke into his apartment.
She is now hiding from the police.

is normally inappropriate, if the indefinite noun phrase *a woman* has not wide scope with respect to *believe*.

In chapter 1, I argued that belief states should be modelled by sets of possible worlds. Now we have seen that we can account for intentional identity attributions in this way, if we are willing to accept the existence of counterpart relations between (i) objects in different worlds, and (ii) belief objects of different individuals.

3.9 Speaker's reference and common grounds

In this chapter I have not yet discussed the *variable aboutness problem about attitudes de re* related with case 4. The problem was that for case 4, (7) is true while (8) is false on their most straightforward readings:

(7) Someone murdered Smith, and Arsky thinks he didn't murder Jones.
(8) Someone murdered Smith, and Arsky thinks he murdered Jones.

The difference between (7) and (8) can be accounted for by the approach defended in the last two sections, where I assumed that the context of interpretation after the first conjunct of (7) and (8) is such that the only relevant counterpart functions are ones that give Arsky's belief object represented by *murderer of Smith* when applied to the actual murderer of Smith and Arsky's belief state. This does not seem unnatural, but it is slightly suspicious. Why should the context select a (set of) counterpart function(s), instead of a salient guise under which the *speaker* thinks about the actual person who murdered Smith, namely *the one who murdered Smith*?

It appears that there is a natural way to account for the asymmetry between (7) and (8). We just assume that (i) *de re* belief attributions should be accounted for in the same way as in chapter 1, and (ii) that pronouns can either be referentially or descriptively used. Then we might say that both (7) and (8) are false on the *de re* reading, because Arsky has no beliefs *about* the actual murderer of Smith, but that (7) is true and (8) false if the pronoun *he* is descriptively used.

On this proposal, we account for the assymetry between (7) and (8) by the use of descriptive pronouns, and for the asymmetry between (5) and (6) by the machinery develloped in sections 3.7 and 3.8.

However, it is unclear why the two asymmetries should be explained in different ways, and, more seriously, the proposals in the last two sections to account for the asymmetry between (5) and (6) are not completely satisfactory. The accounts explain why, on their most natural readings, (5) is true but not (6). However, it seems necessary to explain why (6) is *false* on its most natural reading, rather than merely not unambiguously true.

---

[205] This definition is due to Ede Zimmermann. It is a (± three lines) shorter version of an equivalent formulation I came up with. According to the definition, if there is a $d \in D(w)$ such that for a $c \in C_{acq}(a,w)$ and all $w' \in K(a,w)$: $c_{w'}(d) \neq *$, then there is an $f \in BO(a,w)$ such that for all $w' \in K(a,w)$: $f(w') = c_{w'}(d)$. In the terminology of Kamp (1990) we might say, no external anchor without internal anchor.

In all approaches discussed in this chapter, I assumed that pronouns should either be treated as abbreviations for the antecedent clause or as variables bound by a dynamic existential quantifier. Of course, this is not really the position I favour. As should be clear from chapter 2, I believe that most pronouns are referential expressions that refer to the speaker's referent of the antecedent indefinite. I argued that with the *diagonalisation* strategy of Stalnaker (1978) and the claim that the determination of the referent of the unbound pronoun in the consequent of donkey sentences depends on *counterfactual*, rather than actual, reference-contexts, such an analysis can be pushed further than many have supposed. What is important is that there is an essential difference between the analysis proposed there and standard dynamic semantics. In standard dynamic semantics, the discourse *A man is walking in the park. He is whistling* is said to be truth conditionally equivalent to *A man who is walking in the park is whistling*, whereas this is not true in the semantic analysis I proposed. If the speaker, when he utters the sentence, has a specific man in mind who is walking in the park and this man does not whistle, while another man walking in the park does, the second sentence in which the pronoun occurs is predicted to be false, although *A man who is walking in the park is whistling* will still be counted as true. Observe that on such an account the second sentences of the following two discourses can have different truth values:

(20) A man who is walking in the park wears blue shoes. He is wistling.
(21) A man who is wearing blue shoes is wistling. He is walking in the park

Thus, the two discourses are allowed to be *asymmetric*, something that is not possible according to the dynamic accounts discussed in this chapter. This suggests, of course, that the asymmetry between (7) and (8) should also be accounted for by taking the notion of speaker's referent more seriously than it is in standard dynamic semantics.

In section 2.4 we have given a formal semantics in which the notion of speaker's reference is taken serious. In that section we assumed that indefinites are represented as epsilon terms and assumed that the referent of an indefinite term is determined by means of the choice function of the possibility. Of course, to account for the notion of speaker's reference, there is no need to represent indefinites as epsilon terms. On the assumption that possibilities are represented by quadruples like $\langle\varphi,g,c,w\rangle$, where φ is a choice function, g an assignment function, *c* a counterpartfunction, and *w* a world, we might also interpret formulae of the form '$\exists$xA' in the following way:

$$[[\exists xA]](S) = \{\langle\varphi',h,c,w\rangle \mid \exists\varphi,g \subseteq h \;\&\; \langle\varphi,g,c,w\rangle \in S \;\&\; \langle\varphi',h,c,w\rangle \in [[A]](\{\langle\psi',l,c,w'\rangle \mid \exists\psi,k: \langle\psi,k,c,w'\rangle \in S \;\&\; k[x]l \;\&\; l(x) = \psi(D) = \varphi(D) \;\&\; \psi'(D) = \varphi(D - \{\varphi(D)\})\})\}$$

Note that also in this interpretation rule for indefinites a semantic distinction is made between bound and unbound pronouns. For simplicity, however, I will neglect the difficulties involving epistemic *might*, and will do as if the choice functions won't change after the interpretation of the indefinite. Then we can simplify the above interpretation rule as follows:

$$[[\exists xA]](S) = [[A]](\{\langle\varphi,h,c,w\rangle \mid \exists g: \langle\varphi,g,c,w\rangle \in S \;\&\; g[x]h \;\&\; h(x) = \varphi(D)\}.$$

On this assumption we can interpret belief attributions as follows:

$$[[Bel(a, A)]](S) = \{\langle\varphi,h,c,w\rangle \mid \exists g: \langle\varphi,g,c,w\rangle \in S \;\&\; \exists c' \in C(\mathbf{a},w): \forall w' \in K(\mathbf{a},w): \langle\varphi,h,c',w'\rangle \in [[A]](\{\langle\phi,g,c',w''\rangle \mid w'' \in K(\mathbf{a},w)\})\}$$

But although we have made sense of speaker's reference by this interpretation rule of '$\exists$xA', we still cannot account for the differences between (5) and (6), and (7) and (8), respectively. To account for those differences, I wish to propose, the argument of the choice function should not be the whole set of individuals, but a set of belief objects, and what should always be introduced by an indefinite should be a particular belief object. So, not only in belief attributions will indefinites introduce belief objects, but also in ordinary extensional contexts. But what is then the set of belief objects between which the choice function chooses one? That, I wish to propose, is contextually given, and this contextual ambiguity will formally be represented by anaphoric means. From now on I will use a two sorted language, where the variables *x* and *y* range over individual concepts, and the variables *i* and *j* over information states. I will assume that information states are represented by pairs like <C, K>, where K is a set of possible worlds, and C is a set of counterpartfunctions from individuals and worlds in K to individuals. I will also assume that if <C, K> is, for instance, the belief state of *a* in *w*, we can determine the set of belief objects of *a* in *w*, BO(a,w), just as in § 3.8. I will assume that if <C, K> represents an information state, BO(<C, K>,w) gives us the set of individual concepts that can be defined in terms of C, K and *w*. From now on I will represent 'existential' sentences by formulae of the form '$\exists x_n A$', and interpret them as follows:

$$[[\exists x_i A]](S) = \quad [[A]](\{<\varphi,h,c,w> \mid \exists g: <\varphi,g,c,w> \in S \;\&\; g[x]h \;\&\; h(x) = \varphi(BO(g(i),w)))\}$$

Because assignments assign not objects, but either individual concepts or information states to variables, we have to re-interpret variables as follows:

$$\begin{aligned} \|x\|^{\varphi,g,c,w} &= c_w(g(x)(w)), \text{ if x is of type individual concept and } g(x)(w) \text{ is defined,} \\ &= g(x), \text{ if x ranges over information states} \\ &= \text{undefined otherwise} \end{aligned}$$

Note that although we assign individual concepts to variables, I still use a counterpartfunction, because I will assume that it is possible that we assign to a variable like *i* the following function from worlds to belief states: $\{<w, <C, \{w\}>> \mid w \in W\}$.

I section 1.14 I argued that what is expressed by a sentence depends on the beliefs and intentions of the speaker. I assumed that if the speaker uses a referential expression, he wants to express his beliefs concerning a particular belief object. In that case, the term refers to the causal origin of this belief object. In chapter 2 I argued that also indefinites are referential expressions. That is, also by his use of an indefinite the speaker normally wants to express his beliefs concerning one of his particular belief objects. So, I assume that normally the set of belief objects relevant for the interpretation of a formula like '$\exists x_i A$' in possibility $<\varphi,g,c,w>$ for speaker *a* in *w* will be the following set:

$$\begin{aligned} BO(g(i),w)) = \quad & \{f \in [K(a,w) \rightarrow D] \cup \{<w,d>\} \mid \exists c \in C(a,w) \;\&\; \forall w' \in K(a,w): \\ & c_{w'}(d) \neq * \;\&\; f = \lambda w'.c_{w'}(d) \,\text{I}\!\!\text{N}\, K(a,w)\}, \\ & \text{where } g(i) = <C(a,w),K(a,w)> \end{aligned}$$

Note that even if *w* is no element of K(a,w), a function in BO(g(i),w)) will still assign a value to *w* if the corresponding function in [K(a,w) $\rightarrow$ D] has a causal origin. In effect, if $w \notin K(a,w)$, *f* is an element of [K(a,w) $\rightarrow$ D] and $f' = f \cup \{<w,d>\}$, then $f'(w)$ will be the individual the belief object *f* is about. That is, *d* will be the individual that is represented in <C(a,w), K(a,w)> by *f*.

To account for the asymmetries between (5) and (6), and (7) and (8), respectively, we have to assume that indefinites do not always introduce belief objects of the speaker himself. In our particular examples they should rather introduce 'belief' objects of the common grounds between Arsky and Barsky, and the speaker and Arsky, respectively. This can be accounted for if we assume that also common grounds between two or more individuals should be represented by pairs like <C, K>, and because we have made it contextually (anaphorically) dependent what the relevant information state is from which a 'belief' object is chosen by the speaker by his use of the indefinite.

Why can we in this way account for the asymmetry problem? Consider first (7) and (8), represented by (7') and (8') respectively:

(7') $\exists x_i MS(x).\ Bel(a, \sim MJ(x))$

(8') $\exists x_i MS(x).\ Bel(a, MJ(x))$[206]

To account for the asymmetry problem I need to make two reasonable assumptions. First, I need to assume that the variable *i* will refer for both examples to the common ground between the speaker and Arsky.[207] Because it is presupposed that Arsky has a two-murderer theory, this common ground will contain at least two distinct 'belief' objects that actually refer to the same individual. The second reasonable assumption I have to make is that in both examples the 'belief' object introduced under *x* is the function that assigns the one who shot Smith to each world compatible with what Arsky believes.

The asymmetry between (5) and (6) can be explained in almost the same way. The only difference seems to be that the variable *i* should not refer to the common ground between the speaker and Arsky, but the common ground between Arsky and Barksy. In general, to account for the twin-earth version of Edelberg's Astronomers, we will assume that if more agents are relevant, the relevant information state will always be either an information state due to conversation between the two agents, or an information state construable out of the belief states of the agents which contains only belief objects construable out of specific belief objects of the two agent's belief states that each have the same source.

3.10 Conclusion

In this chapter I have discussed several ways in which we might account for intentional identity attributions. On the assumption that unbound pronouns are either treated as abbreviations for the antecedent clause, or as variables bound by a dynamic existential quantifier, I could not come up with a completely satisfactory analysis. This, or course, can be no impossibility proof for such an analysis. Perhaps I just didn't see it. This, however, is not the way I prefer to think of it. I would like to think that the phenomena discussed in this chapter show the limits of the received doctrine with respect to the analysis of so-called unbound pronouns, and that it is better to make speaker's reference *truth-conditionally* relevant. This conclusion I share with Edelberg (1992, 1995).

Still, I don't follow Edelberg's main conclusion. Edelberg argues that intentional identity attributions suggest a perspectival account of semantics. The account proposed in section 3.10 on the other hand is realistic, in the sense that what is expressed by a sentence is true or false in a world, and not just with respect to a theory or belief state. I have argued that a realist analysis of the relevant phenomena is both empirically preferred, and no more complicated than the perspectivalist analysis proposed by Edelberg.

---

[206] Note that I represent (7) and (8) as discourses, because only then speaker's reference becomes relevant.

[207] Or at least that *n* refers to a function from worlds to the belief state of Arsky in this world extended with this world itself.

In this chapter, I have made essential use of entities called *belief objects*. In whatever way these belief objects are modelled, by using them we offer a substantial hypothesis: namely that belief states cannot be modelled simply by a set of possible worlds, if these worlds, and the relations between those worlds, can be characterised completely by purely qualitative or descriptive means. Indeed, I believe that one of the main insights of dynamic semantics is that it shows, once again, that information states cannot be characterised in terms of purely general concepts.

In this chapter we have almost only looked at the problem of intentional identity in case of belief attributions. Similar phenomena involving presuppositional dependencies across belief attributions have almost not been discussed. In the next chapter I will give some attention to this problem. In this chapter I also neglected intentional identity cases where other attitudes than *belief* are involved These phenomena have a lot in common with the analysis of intentional identity attributions, but there are, I will argue, also some essential differences. Assuming the existence of *belief objects*, for instance, is useful to account for the analysis anaphoric relations across belief attributions, but does not seem to help in case of *desire* attributions. In order to account for anaphoric dependencies across desire attributions, I will argue that *belief revision* must be taken into account. In chapter 5 I will discuss some analyses of belief revision, and in the last chapter of this dissertation I try to use this for the interpretation of some other attitude attributions.

Appendix

To prove $X \vdash_d A$ iff $W(X) \models_d A$

Proof:
The 'hard' part is to prove that $f_w(A, g) \neq \varnothing$ iff $[[A]]\ \{<g,w>\} \neq \varnothing$. Once that is done, the proof is simple:

(1) $X \vdash_d A$ iff $\forall <g,s> \in X$: $\forall w \geq s$: $f_w(A, g) \neq \varnothing$ iff $\forall <g,w> \in W(X)$: $f_w(A, g) \neq \varnothing$

(2) $W(X) \models_d A$ iff $\forall <g,w> \in W(X)$: $[[A]]\ \{<g,w>\} \neq \varnothing$

(3) $f_w(A, g) \neq \varnothing$ iff $[[A]]\ \{<g,w>\} \neq \varnothing$

(4) $X \vdash_d A$ iff $W(X) \models_d A$ (by (1), (2) and (3))

Proof $f_w(A, g) \neq \varnothing$ iff $[[A]]\ (\{<g,w>\}) \neq \varnothing$ by induction on complexity of formula

Base step:

(1a) $f_w(Px_1,..., x_n, g) \neq \varnothing$ iff $<g(x_1),..., g(x_n)> \in I_w(P)$ iff $[[Px_1,..., x_n]]\ (\{<g,w>\}) \neq \varnothing$

(1b) $f_w(x = y, g) \neq \varnothing$ iff $g(x) = g(y)$ iff $[[x = y]]\ (\{<g,w>\}) \neq \varnothing$

Induction Hypothesis: $f_w(C, g) \neq \varnothing$ iff $[[C]]\ (\{<g,w>\}) \neq \varnothing$, which is equivalent to
$f_w(C, g) = \varnothing$ iff $[[C]]\ (\{<g,w>\}) = \varnothing$, where C is A or B.

(2) $f_w(\sim A, g) \neq \varnothing$ iff $f_w(A, g) = \varnothing$ iff (IH) $[[A]]\ (\{<g,w>\}) = \varnothing$ iff $\sim\exists h[g \leq h$ & $<h,w> \in [[A]]\ (\{<g,w>\})]$ iff $[[\sim A]]\ (\{<g,w>\}) \neq \varnothing$.

(3) $f_w(A \wedge B, g) \neq \varnothing$ iff $\cup \{f_w(B, h) : <h,w> \in f_w(A, g)\} \neq \varnothing$ iff (2 × IH)
$\cup \{[[B]]\ (\{<h,w>\}) : <h,w> \in [[A]]\ (\{<g,w>\})\} \neq \varnothing$ iff
$[[B]]\ ([[A]]\ (\{<g,w>\})) \neq \varnothing$

(4) $f_s(A \vee B, g) \neq \varnothing$ iff $f_s(A, g) \neq \varnothing$ or $f_s(B, g) \neq \varnothing$ iff (IH)
$[[A]]\ (\{<g,w>\}) \neq \varnothing$ or $[[B]]\ (\{<g,w>\}) \neq \varnothing$ iff $[[A \vee B]]\ (\{<g,w>\}) \neq \varnothing$

(5) $f_s(\exists xA, g) \neq \varnothing$ iff $\cup \{f_s(A, g[^x/_d]) | d \in D\} \neq \varnothing$ iff (IH)
$\cup \{[[A]]\ (\{<g[^x/_d],w>\}) | d \in D\} \neq \varnothing$ iff $[[\exists xA]]\ (\{<g,w>\}) \neq \varnothing$

This completes the proof.

Chapter 4

Presuppositions and two dimensions

4.1 Introduction

In the foregoing chapters we have assumed that presupposition is a propositional attitude and that what is presupposed is what the speaker takes to be the presumed common knowledge between speaker and hearer. As a result, presuppositions should be given a *pragmatic* analysis: what a *sentence* presupposes should be explained in terms of what *speakers* normally presuppose by their use of these sentences. This analysis will be used in this chapter to account for so-called presuppositional inferences from the use of certain sentences. For instance, to account for the fact that we are normally allowed to infer from the speakers assertion of both *John regrets that he failed* and *john doesn't regret that he failed* that John failed.

In the seventies, truth-conditional semantics was considered to be a self-contained subject matter, something that can be studied in abstraction from pragmatics. It was argued that what is normally presupposed by the use of a certain sentence *can* be separated from the *content*, its truth-conditions, of what is said by the sentence. Moreover, it was considered to be *useful* to separate presuppositions from truth-conditions, because in that case one could also separate the question of entailment from that of presupposition, and thereby simplify a semantic theory. This view gave rise to the *two-dimensional* approaches towards presuppositions, according to which what is asserted by a sentence can be determined independently of what is presupposed. The best known two-dimensional analyses of presuppositions has been developed by Gazdar (1979) and by Karttunen & Peters (1979). They assumed that what is asserted and what is presupposed can be represented by separate propositions, and that what is asserted is independent of what is presupposed. The two-dimensional accounts have been abandoned by almost everybody since the eighties. This was due to both the great attention the problem of context-dependence has received in the dynamic theories of meaning, and to Karttunen & Peters' (1979) *binding problem*. Both gave rise to a similar conclusion: representing what is presupposed and what is asserted by separate propositions neglects the dependency of what is asserted on what is presupposed.

In this chapter I will make and defend two claims: (i) the binding problem and the similarity between anaphora and presuppositions do not show that the two-dimensional account is mistaken, they only show that presupposed information is not only information about the subject matter of conversation, and that the separate calculations of what is asserted and what is presupposed cannot give rise to meanings that represent truth-conditional contents. These meanings must be more abstract entities such that the truth-conditional content of what is asserted can depend on what is presupposed. (ii) a two-dimensional analysis of presuppositions is not only possible, it is also desired, because it allows for a separation of semantic entailment and presuppositional inference, and, unlike other approaches, a two-dimensional analysis can very naturally account for the presuppositions of quantified sentences.

The two-dimensional approaches towards presuppositions are closely related to the two-dimensional theory of reference discussed in chapters 1 and 2. The two-dimensional theory of reference is based on the assumption that we can separate facts about the subject matter of conversation from certain facts about the conversational situation itself. The two-dimensional theories of presuppositions are based on the assumption that we can separate the truth-conditions of sentences from the constraints these sentences put on the contexts in which these sentences can be appropriately uttered. Still, it has been commonly assumed that those two theories were independent of each other. But this, obviously, cannot be the case if it is assumed, as we do, that one of the most important fact about a conversational situation is what the speaker presupposes. On the other hand, what the speaker

presupposes is related both to facts about the conversational situation, and to facts about the subject matter of this conversation.

This chapter is roughly divided into three parts. The first part consists of § 4.2 until § 4.6. In this part I argue for and formulate a two-dimensional approach that can be thought of as a combination of the approaches of Gazdar (1979) and Karttunen & Peters (1979). In the second part, I discuss how to account for presuppositions in attitude contexts. In the most important part of this chapter I will propose a two-dimensional treatment of so-called *anaphoric presuppositions*, and of presuppositions in *quantified contexts*. With respect to the former, I argue throughout the whole chapter that presuppositional inferences cannot be just about assumed background information about the *subject matter* of the conversation, presuppositions can also be about background assumptions of the *conversation itself*. In the final part of this chapter I argue that the use of so-called anaphoric presupposition triggers like *too* and *also* presupposes (normally) that a certain kind of expression has already been used in the discourse, and account for this in a two-dimensional way. With respect to the latter, I defend the claim that quantified sentences should give rise to presuppositions that select the domain of quantification for the assertion. I argue that, contrary to other approaches, this suggestion can be accounted for naturally in a two-dimensional framework. The binding problem of Karttunen & Peters (1979), I will argue, is just the result of a too coarse-grained analysis of presuppositions. What is asserted and what is presupposed need not be a proposition, but should be thought of as a context change potential, or ccp. The ccp associated with what is asserted is applied to the ccp associated with what is presupposed. As a result, what is asserted can be determined independently of what is presupposed, although the proposition expressed cannot. Assuming that we determine a ccp associated with what is asserted and what is presupposed, I will propose that a sentence like *A man is sick* gives rise to the existential presupposition that the set of men is non-empty. The assertion is then about one element of this set. Finally, I formalise this proposal in the CCT framework.

4.2 Presuppositions as definedness conditions

In chapter 2 we saw that in Context Change Theory a context should be represented by a set of reference-context/index pairs. In chapter 1 we have discussed the Kaplanian insight that the meaning of a sentence is a rule for determining the proposition it expresses in different reference-contexts. Combining both, all we need to assume is that a proposition is expressed by an assertive utterance in all reference-context of the possibilities of the context. If a referential expression like a pronoun or a short definite description is used in an utterance, as in *He is sick*, the proposition expressed, if any, by the utterance depends normally on the reference-context. A pronoun or short description normally refers to the unique most salient individual that is presupposed to satisfy the descriptive content of the noun phrase. Of course, the hearer need not know what the actual speaker's referent of the noun phrase is, but the speaker assumes that his use of the description does not give rise to new referential ambiguity. Thus, there need not be a unique individual that is presupposed to satisfy the descriptive content of the noun phrase, but it is only presupposed that there is a unique individual that satisfies this descriptive content. The reason is obvious: if it is not clear, for each possibility of the context, what object is referred to by the definite term, it is not clear what, if any, (horizontal) proposition would be expressed by the assertion if this possibility would be the actual reference-context/index pair, and thus it need not be clear for the hearer whether the statement made by the assertion is true or false in the world of that possibility. The assertion can neither be characterised as true nor as false in this possibility. If for some possibilities in the context what is said is neither true nor false, it follows that it is not clear for the participants of the conversation how the context should be updated, something that should be avoided. Considerations such as these gave rise to the Frege (1892)/Strawson (1950) notion of *semantic*, or *expressive, presupposition*. An utterance of a sentence like *The S is P* presupposes that it is clear, for each possibility in the context, which *S* it refers to. In chapter 2 we characterised truth as follows:

formula A is *true in* <g,h,w>, <g,h,w>⊨A, iff {<g,h,w>}⊨$_d$ A

Let us say for simplicity that triples like <g,h,w> are simply worlds. Let us also assume that for every v ∈ W: v(A) = 1 if and only if v⊨A, and v(A) = 0 if and only if v⊨~A. If we not only restrict ourselves to possibilities of the context, the notion of 'semantic presupposition' gives rise to a two-place logical relation between sentences or propositions:

B is a *presupposition* of A iff for all v ∈ W: if v(B) ≠ 1, then v(A) ≠ 1 and v(A) ≠ 0

In worlds in which there are more than one *S*, but no most salient *S*, a sentence of the form *The S is P* can neither be characterised as false, nor as true, because the presupposition of the sentence is not satisfied in this world. The sentence will have a third truth-value. Thus, according to a three-valued account, if B is a presupposition of A, and v(A) ∈ {0,1}, it follows that v(B) = 1. For instance, if A is *The man is sick* and v(A)∈ {0,1}, it follows that in *v* there is a unique most salient man available for reference.

Given that a simple sentence like *The S is P* gives rise to the above presupposition, the question arises what is presupposed by a sentence of which this simple clause is a part. For the negation of the original sentence the answer seems straightforward. A sentence is false in a possibility, if its negation would be true in this possibility.

(1) *The man* is not sick.

Thus, a sentence of the form ~A presupposes the same as A itself. So much for negation. What about the other connectives? The simplest hypothesis would be that what seems the case for negated sentences also holds for these other constructions. Thus, what is presupposed by a complex sentence is simply the sum of what is presupposed by its parts.[208] On the basis of the assumption that all definite noun phrases are used in the same way, this comes down to the hypothesis that for every definite term of the form *The S* used in a complex sentence, it is presupposed that there exists a unique most salient *S* for the sentence to be used appropriately. However, Russell (1905) noted already that this hypothesis is false. If a definite term is used in the consequent of a conditional whose antecedent gives us enough information to determine the referent of this definite term, the appropriateness condition required for the consequent will not be a condition for appropriate use of the whole conditional. In modern terminology, the presupposition of the consequent can be *cancelled.* For other constructions similar observations have been made; the presupposition of the second conjunct of a conjunction can be cancelled if this presupposition follows from what is asserted by the first conjunct, and the presupposition of the second disjunct of a disjunction can be cancelled if this presupposition follows from the negation of the first disjunct.

(2a) If *the S* is walking in the park, it is not raining.
(2b) If an S is walking in the park, *the S* is walking his dog.
(3a) *The S* is walking in the park, and it is not raining
(3b) An S is walking in the park, and *the S* is walking his dog
(4a) *The S* is on the second floor, or on the third.
(4b) Either there is no S in the house, or *the S* is on the second floor.

Let us assume that what is presupposed by a clause that is negated, that figures as the antecedent of a conditional, or that figures as the first conjunct or disjunct of, respectively, a conjunction and disjunction, is the same as what is presupposed by the clause itself. Let us moreover assume that what is presupposed by a clause figuring as the consequent of a

[208] This is assumed by Kaplan (1989) for what he considered to be referential expressions. We assume, however, that also anaphoric pronouns are normally referentially used, and thus we cannot take over Kaplan's assumption.

conditional, or as second conjunct or disjunct, depends on whether or not what is presupposed by the clause itself is entailed by, respectively, the antecedent of the conditional, by the first conjunct, or by the negation of the first disjunct. In that case the behaviour of presuppositional inferences can be accounted for in a truth-conditional way by the use of the following truth-tables due to Peters (1977):[209]

	$\neg A$	$A \wedge B$	$A \rightarrow B$	$A \vee_a B$
		B		
A				
1	0	1 0 *	1 0 *	1 1 1
0	1	0 0 0	1 1 1	1 0 *
*	*	* * *	* * *	* * *

To be able to determine *the* presupposition of a sentence, let us define a definedness operator, $\Downarrow$, a function from formulae to valuation functions, $\Downarrow(A) = \{v \in V \mid v(A) \in \{0,1\}\}$ (where V is a set of valuation functions, and A an atomic formula).

By the truth tables given above, we can derive the following definedness (or satisfaction) conditions:

$$\Downarrow(\neg A) = \Downarrow(A),$$
$$\Downarrow(A \wedge B) = \Downarrow(A \rightarrow B) = \Downarrow A \cap \{v \in V \mid v \in [A]^+ \rightarrow v \in \Downarrow B\},$$
$$\Downarrow(A \vee_a B) = \Downarrow A \cap \{v \in V \mid v \in [\neg A]^+ \rightarrow v \in \Downarrow B\},$$

where $[A]^+ = \{v \in V \mid v(A) = 1\}$.

By means of the above truth-tables we can account for the above assumed presuppositional behaviour of complex sentences build up by means of negation, implication, conjunction, and disjunction in a purely truth-conditional way. A truth-conditional account of presuppositions suggests that presuppositions can be accounted for independently of the conversational situation in which the presupposition triggers are used. But this gives rise to two related worries. First, it seems impossible to satisfy the presupposition associated with (1) that there is a unique most salient man available for reference by a short description in a discourse-independent way.[210] Second, it is not clear whether an *independent* motivation can be given for the above definition of the connectives. For the case of *negation* such a motivation has been given by Strawson (1950) in terms of truth-conditions only. But what about the other connectives? Because in Peters' logic conjunction, implication and disjunction are defined in an asymmetric way, the question is how the asymmetric meaning of these connectives can be argued for. A purely *truth-conditional* motivation is hard to imagine.

With respect to the first worry, we have assumed that presuppositions have to be satisfied by the context that represents the background information about the subject matter, *and* the discourse. Stalnaker (1973, 1974, 1978) and Karttunen (1974) explained the asymmetric behaviour of presuppositions in compound sentences, the second worry, in terms of the way sentences *change* the background assumptions in the course of conversation. The context represents the information assumed (for the sake of conversation) by the

---

[209] In fact, Peters reconstructed Karttunen's (1974) satisfaction-theory in a three-valued logic.

[210] See already Russell (1957) who claimed that Strawson (1950) confused the problem of description with that of egocentricity.

participants of a conversation about both the subject matter of conversation, and the conversation itself. This context of interpretation changes, if new assertions are made, and if the participants accept new information. If an assertion is made and accepted, the new context might contain other objects available for reference by pronouns or short descriptions, and will be such that each possibility of the context verifies what is expressed by an utterance in that possibility. Background assumptions do not only change after a whole sentence is asserted. In CCT it is assumed that, in case of a conjunction, the assumptions change in the middle of the assertion. The context of interpretation for the second conjunct will be the initial context incremented by the information available from the first conjunct. So, there is no constraint on the initial context for the assertive utterance of (3b) because by the above assumption, the context of interpretation for the second conjunct will satisfy its constraint. The same story holds for (2b), if we make the natural assumption that the consequent is interpreted in a context in which all the information available from the antecedent is assumed.

We have seen that in order to explain presuppositional phenomena, we should not look primarily at the truth-conditions of sentences, but at the way the assertion of sentences *changes* its context of interpretation. Let's therefore quickly give a propositional version of CCT as it was used by Stalnaker (1973, 1974) and Karttunen (1974), with the added assumption that presupposition failure gives rise to undefinedness. The semantics is defined with respect to a model $\langle W, \models \rangle$, where W is a set of worlds, and '$\models$' a satisfaction relation between worlds and atomic formulae. Let us suppose that if A is an atomic clause, $A_{\langle P \rangle}$ means that P is the presupposition associated with A. The ccp of sentence A, [[A]], can now be recursively defined as follows:

$[[A_{\langle P \rangle}]](s) = \{w \in s \mid w \models A\}$, if A is atomic, and $\forall w \in s: w \models P$
undefined otherwise

$[[\neg A]](s) = s - [[A]](s)$

$[[A \wedge B]](s) = [[B]]([[A]](s))$

Disjunction and implication can be treated syncategorematically, by having '$(A \vee_a B)$' and '$(A \rightarrow B)$' stand for '$\neg(\neg A \wedge \neg B)$' and '$\neg(A \wedge \neg B)$' respectively.

We say that A entails B, $A \models B$, if and only if for every model and every $s \subseteq W$: $[[A]](s) = [[B]]([[A]](s))$.

On the basis of these context-change rules, we can immediately calculate the presupposition projection rules. Let us use '$\downarrow$' for the definedness function from sentences to propositions: that is, $w \in \downarrow A$, if A atomic, if and only if $w \in \cup dom([[A]])$. It assigns to every sentence the set of possibilities in which it is defined. For atomic clauses the definedness-function assigns to it the set of possibilities in which its presupposition is true.[211] Let us define the characteristic set of A, c(A), as follows:

$c(A) = \{w \in W \mid [[A]](\{w\}) = \{w\}\}$

Then, as result of the ccp-conditions, for negations, conjunctions, conditionals and disjunctions the following follows:

$\downarrow(\neg A) = \downarrow(A)$

$\downarrow(A \wedge B) = \downarrow(A \rightarrow B) = \downarrow A \cap \{w \in W \mid w \in c(A) \rightarrow w \in \downarrow B\}$

---

[211] It is important to realise that on this account every sentence has only one presupposition.

$$\downarrow(A \vee_a B) \quad = \quad \downarrow A \cap \{w \in W \mid w \in c(\neg A) \rightarrow w \in \downarrow B\}$$

This shows the great similarity between the presuppositional predictions made by the satisfaction approach towards presuppositions in context-change theories, and the three-valued logic stated above.[212] The similarity appears less accidental, if it can be shown that the sets $[A]^+$ and c(A) are very closely connected with each other. To see the connection, it is shown in the appendix of Van Rooy (1994) that the models of Peters' three-valued logic's can be translated into a context change model and vice versa such that the same presuppositional predictions are made.[213] Still, the similarity is somewhat misleading. Some triggers presuppose something about the discourse itself, not about the subject matter of the discourse. Moreover, the information about the discourse itself changes constantly during the discourse. This can be accounted for in a straightforward way in Context Change Theory, but it is not clear how to account for this in a purely truth-conditional approach.

4.3 Some problems with presuppositions as definedness conditions

Until now we have tried to explain two things about presuppositions. First, why do simple sentences have the presuppositions they have? Second, given that simple sentences have the presuppositions they have, why do compound sentences have the presuppositions they have? The first question was answered in terms of definedness conditions. Sentence A is said to presuppose B, if without the assumption that B, it cannot be assumed that a unique proposition is expressed by A.[214] In the way we model the background assumptions, this means that there are some possibilities of the context in which either *no* horizontal proposition is expressed by A, or with respect to which it cannot be determined *what* horizontal proposition is expressed by sentence A. If in some possibilities of the context it cannot be determined whether what is expressed by A in that possibility is also true in that possibility, it is not clear to the hearer whether he should eliminate this possibility from the context or not. This, however, is in conflict with one of the major Gricean maxims about coöperative conversation. For the second question we pointed at the way the background assumptions change during a conversation. We supposed that the context of interpretation for the second conjunct, or the consequent of a conditional, was the initial context updated by the first conjunct or the antecedent.

But there are well known problems especially with the first answer; the explanation of why simple sentences have the presuppositions they have. This explanation is built on the assumption that when the presupposition of a sentence does not hold in a possibility, it cannot be determined what proposition, if any, is expressed in this possibility. This might be a natural way to think of sentences like *The S is P*, where *the S* is a referentially used description or pronoun, but seems to lose its force for other constructions that intuitively give rise to inappropriate speech acts if certain conditions are not satisfied, and moreover show the same projection behaviour as constructions for which 'presupposition as definedness condition' makes sense. For instance, although both *The Queen of Holland has three sons* and *The Queen of Holland doesn't have three sons* give rise to the inference that there is a unique Queen of Holland, there seems to be no convincing reason why these sentences would not express a determinate proposition with respect to reference-contexts that do not contain a Queen of Holland. The reason is that descriptions need not be referentially used, they can be *attributively used*, too. If a description is attributively used, the speaker does not intend that the truth or falsity of the sentence in which the description

---

[212] Other truth-functional three-valued logic's correspond with other ccp-rules. Note that Van Fraassen's (1966) supervaluation account is not truth-functional.

[213] The first who saw the exact connection between presupposition_as_definedness-condition in a context-change theory and partial semantics is Peters (1977).

[214] Which need not be the same as: there is a unique (horizontal) proposition that can be assumed to be expressed.

occurs is a function of the person or thing the speaker has in mind. The rule for determining the referent of the description is part of the (horizontal) proposition expressed. Russell (1905) and others have argued that when this rule is part of the proposition expressed, the sentence containing the description that gives rise to this rule is simply false in a world in which the rule does not determine a unique referent. Something similar holds for factive verbs. Although we normally infer from both *John knows that A* and *John doesn't know that A* that *A* is the case, it is not at all clear that no proposition is expressed by the sentence *John knows that the earth is flat*, although in fact the earth is not flat. Why not say that the latter sentence expresses a proposition that is simply false in the actual world? For other so-called presupposition triggers it has been argued that an atomic sentence in which this trigger occurs can even be true if the associated presupposition does not hold. It has been argued, convincingly, in Horn (1969), Stalnaker (1973) and Karttunen & Peters (1979) that a word like *even* makes no contribution to what is asserted; what is asserted by the sentence *Even John came* has the same truth-conditional content as what is asserted by *John came*. As a consequence, the truth-conditions of the former can be determined independently of the presupposition associated with the sentence. The same holds for subjunctive conditionals (Stalnaker, 1973). The use of such a construction invites the inference that its antecedent is false, but there is no good reason to assume that the truth-value of the conditional cannot be determined in a world in which the antecedent is not false. A bit more controversially, a clause like *Some S is P* conversationally implicates that not all S are P. And again, this kind of implicature shows the same projection behaviour as those constructions that give rise to constraints that naturally can be thought of as definedness conditions.

There are several possible answers to give for a proponent of semantic presupposition. First, it can be claimed that those constructions that give rise to constraints that cannot so naturally be thought of as definedness conditions should not be considered as presupposition triggers at all. You might say, for instance, that presupposition triggers are really referential or anaphoric expressions, and that the other constructions should be thought of as something like conversational implicatures. That's a legitimate move and there is certainly something appealing to it, too.[215] Semantic presuppositions give rise to *constraints on* the *reference-contexts* of the context of interpretation, while conversational implicatures give rise to *constraints* related to what is presupposed about the *subject matter of conversation*. The constraints these different constructions make on contexts of interpretation are different in kind, and we should not try to treat them all in the same way. Appealing as such a move might be, it has two unwelcome consequences: (i) it becomes unclear how it can be explained that all the considered constructions show roughly the same projection behaviour, and (ii) it is in need of explanation why for almost any referentially used expression the descriptive material of the noun is essential to determine the referent in a possibility.[216] In sum, it seems that not only a straightforward generalisation is lost, but also that most definite noun phrases cannot be treated as presupposition triggers anymore. Another reply might be to claim that sentences in which these other constructions occur also give rise to truth-value gaps if the associated presupposition does not hold. It is certainly legitimate to make such a move, but the claim that presuppositions have to be explained in terms of truth-conditions is empty, if based on the assumption that the facts about presuppositions cannot be separated from a truth-conditional analysis. The claim that the *descriptive* notion of a presupposition coincides with the *theoretical* notion, does not make them coincide. Anyway, if these other constructions are treated as presupposition triggers, the intuitive explanation given by proponents of the semantic notion of presupposition for what a presupposition is, loses its force.

If presuppositions are treated as definedness conditions, we can only treat these other constructions as presupposition triggers by giving up its original motivation; that it

---

[215] See Russell (1957), Karttunen & Peters (1979), and Soames (1989).

[216] cf. the argument of Groenendijk *et al.* (1995b) against the familiarity approach to definites of Heim (1982).

becomes impossible to determine what is asserted if the presupposition is not satisfied. Stalnaker (1973, 1974) argued that instead we should widen the truth-conditional motivation for presupposition. An assertive use of an utterance presupposes not only that a proposition is expressed by the utterance, but also that what is expressed is conversationally appropriate. A sentence has a presupposition not so much because otherwise we cannot determine for each possibility what is expressed with respect to that possibility, but because otherwise the use of the sentence would *for some reason* normally be *inappropriate*. This, of course, does not rule out a truth-conditional analysis of presupposition triggers. Indeed, the most obvious reason for inappropriateness is lack of truth-value.[217] Just like according to the semantic account, if a sentence has a presupposition, the presupposition has to be *satisfied* in its context of interpretation. But if we think of presupposition in such a more general way as conversational appropriateness condition, we can not only treat these other constructions as presupposition triggers for which presupposition as definedness condition doesn't make much sense, we can also account more easily for the two-way interaction between context and presupposition.

According to a pure satisfaction account towards presuppositions, the presuppositions associated with a sentence have to be already commonly assumed by the participants of the conversation according to the speaker. The approach was motivated by the assumption that discourses proceed in an ideal orderly way,[218] that what is presupposed is already common ground. However, as recognised by Stalnaker (1973) and Karttunen (1974), conversation does not always proceed in an ideal orderly fashion.

> as soon as there are established and mutually recognized rules relating what is said to the presumed common beliefs, it becomes possible to exploit those rules by acting as if the shared beliefs were different than they in fact are known to be. The existence and mutual recognition of the rules is what makes it possible to communicate such a pretence, and thus to use the pretence to communicate. Since we want to say that the presuppositions are present even when such rules are being exploited in this way, we cannot simply identify presupposition with what is actually taken to be common knowledge. (Stalnaker, 1973, p. 451)

> The pretence need not be an attempt at deception. It might be tacitly recognized by everyone concerned that this is what is going on, and recognized that everyone else recognized it. In some cases, it is just that it would be indiscreet, or insulting, or tedious, or unnecessarily blunt, or rhetorically less effective to openly assert a proposition that one wants to communicate. (1974, p. 202)

So, Stalnaker (1973, 1974) conceded that speakers can introduce new information indirectly by presupposition for the sake of brevity, convenience or for some special conversational purposes. Karttunen even remarked that

> People do make leaps and shortcuts by using sentences whose presuppositions are not satisfied in the conversational contexts. This is the rule rather than the exception, and we should not base our notion of presupposition on the false premise that it does not or should not happen. (Karttunen, 1974, p. 191)

On the basis of this, Gazdar concluded with respect to the context-satisfaction approach and the way we might try to cope with the problem in this framework:

> These strategies have a number of methodologically undesirable features: They involve treating the bulk of the data (i.e. ordinary conversation) as something special, they circumvent any possibility of counterexamples and, concomitantly, they render the inclusion of a notion like "appropriacy" in the definition wholly vacuous. (Gazdar, 1979, p. 107)

Gazdar suggested that it is better not to think of presuppositions as definedness conditions at all; what is presupposed by the utterance of a sentence is simply a special kind of inference. But it is not clear how we should then account for the presuppositions associated

---

[217] Thus, I don't agree with Soames (1989) who claims that truth value gaps don't explain presuppositions. The notion of presupposition I favour is a *generalisation* of the semantic notion of presupposition.

[218] Itself based on the assumption that the participants of a conversation are coöperative rational agents.

with referentially used expressions. I think it is better to judge Gazdar's conclusion given in the quotation above as being not justified. One can assume that if a speaker makes a presupposition in a context in which it is not already explicitly established that the presupposition is met, he assumes that the other participants of the conversation believe, or are likely to accept, that the presupposition is true. The other participants simply assume that the context was such that it already satisfied the presupposition, because otherwise the assertion would not be appropriate. They conclude that the context contains more information than it seemed to contain, the context was such that it already satisfied the presupposition. This Gricean strategy referred to by Stalnaker in the quotes above was dubbed *accommodation* by Lewis (1979b). Of course, the use of this repair strategy of the satisfaction account is based on the assumption that *normally* the presupposition associated with an utterance is already met in its context of interpretation. But I believe that this is indeed the case, because pronouns should be seen as (the most normal) presupposition triggers.

How can we consistently account for accommodation if we want to stay as closely as possible to the presupposition_as_definedness account? We have seen that the presupposition of a conjunction or an implication is conditional in nature according to the accounts we have discussed above. Thus, if we should accommodate a presupposition of a conjunction, so you might argue, you should also accommodate a conditional presupposition. If we represent the context by *s*, a set of worlds, it seems that accommodation of *s* with the presupposition of *A* results simply in $s \cap \downarrow A$. But now there is a problem. According to the semantic account there is no need to separately calculate the presuppositions of complex sentences: all we need to know is what atomic clauses presuppose. What is presupposed by complex clauses should follow from the interpretation rules. We seem to have a conflict here. Beaver (1992) showed us how to resolve this conflict. To account for accommodation in a pure satisfaction framework we should not represent a context by a set of possible worlds, but rather by a set of sets of possible worlds. Let's call such a set of worlds a *context set*. The speaker knows what he is presupposing. Unfortunately the hearer doesn't. Several context sets are possible. Being cautious, he doesn't want to choose, and represents the context by a *set* of context sets.[219] Let's denote such a context by K. The context change potential for *A* at this level, [A], is then defined as follows

$$[A](K) = \{[[A]](s) \mid s \in K \,\&\, s \in \mathrm{dom}([[A]])\}$$

Note that whereas for an arbitrary A, [[A]] can be a partial function, because undefined for some $s \subseteq W$, [A] will always be a total function, defined for all $K \subseteq \wp(W)$.

4.4 Ambiguity of logical connectives

We tried to take a pragmatic approach that is close enough to the semantic strategy, but rich enough to account for accommodation of context. However, by staying so close to the semantic account, some of its problems are still with us; since the notion of presuppositional requirement still does not come out as a special kind of appropriateness condition between assertion and context of interpretation, there is no natural way to deal with certain cases of presupposition cancellation; moreover, the peculiar way of how presuppositions of disjunctions are calculated appears to be quite unmotivated.

---

[219] You might argue, he diagonalises again. This can be seen if a context is represented by a set of worlds, and when what is presupposed by the speaker is modelled by an accessibility relation, Presup(c,w). Everything that is accepted during the conversation *c* is known by the hearer to be presupposed by the speaker and thus will be accepted in Presup(c,w) for any world *w* of the context. But that still leaves open that for two worlds *w* and *w'* in the context, Presup(c,w) ≠ Presup(c,w'). If a new utterance is made that presupposes *B*, it might be that *B* is accepted in Presup(c,w) but not in Presup(c,w'). In that case, w' is filtered out by presupposition accommodation.

Because of examples like the following:

(5) John doesn't regret failing, because, in fact, he passed.
(6) If I realise later that I have not told the truth,
I will confess it to everyone. (Stalnaker, 1974)
(7) Either all of Jack's letters have been held up,
or he has not written any. (Karttunen, 1973)

any account of presupposition-projection that determines the actual presupposition made by using only Peters' truth-tables, or the ccp rules given above, will give wrong predictions for (5)-(7). A purely truth-conditional account of presuppositions that takes factives, descriptions and partitives to be presupposition-triggers must also postulate the existence of a presupposition cancelling negation, an implication that never preserves the presupposition of its antecedent, or one that makes it dependent on the consequent, and a symmetric disjunction.[220] They can be defined as follows:

	~A	A ⇒ B	A --> B	A $\vee_S$ B
B		1 0 *	1 0 *	1 0 *
A				
1	0	1 0 *	1 0 *	1 1 1
0	1	1 1 1	1 1 1	1 0 *
*	1	1 1 1	1 * *	1 * *

Given these truth-tables and the above definedness (or satisfaction) conditions, we have:

$\Downarrow(\sim A) = V$
$\Downarrow(A \Rightarrow B) = \{v \in V \mid v \in [A]^+ \rightarrow v \in \Downarrow B\}$
$\Downarrow(A \dashrightarrow B) = \{v \in \Downarrow A \cup \Downarrow B \mid (v \in [\neg B]^+ \rightarrow v \in \Downarrow A) \,\&\, (v \in [A]^+ \rightarrow v \in \Downarrow B)\}$
$\Downarrow(A \vee_S B) = \{v \in \Downarrow A \cup \Downarrow B \mid (v \in [\neg B]^+ \rightarrow v \in \Downarrow A) \,\&\, (v \in [\neg A]^+ \rightarrow v \in \Downarrow B)\}$

where $[A]^+ = \{v \in V \mid v(A) = 1\}$.

With the help of the above definition of c(A), we can also define ccp rules for these new connectives:

$[[\sim A]](s) = s - c(A)$,
$[[A \Rightarrow B]](s) = s - (c(A) - [[B]](c(A) \cap s))$,
$[[A \dashrightarrow B]](s) = s - ([[A]](c(\neg B) \cap s) - [[B]](c(A) \cap s))$, and
$[[A \vee_S B]](s) = [[A]]([[\neg B]](s)) \cup [[B]]([[\neg A]](s)) \cup ([[A]](s) \cap [[B]](s))$.

These ccp's have different presupposition-projection properties than the ones given earlier, but give the same truth-conditions if a two-valued logic is assumed (see Soames, 1989).

---

220 Actually, already Bochvar's system (1939) of internal and external connectives predicts an ambiguity of all logical operators. The second implication is known as the strong Kleene implication.

It is easy to see that the three-valued interpretation of the new connectives correspond to new ccp-rules in the same way as the Peters' connectives corresponded with the earlier given ccp-rules.

So, the most straightforward way for proponents of the view that presuppositions are definedness conditions to account for the different readings of *John does not know that Mary failed the test*, and account for the presuppositional readings of (2) -(4) and the non-presupposition preserving (5), (6) and (7), is to postulate *lexical ambiguities* of the connectives.

A number of people (cf. Seuren, 1985) have argued for negation that it is *semantically* ambiguous in this way, but corresponding arguments for disjunction, and implication I have never seen. Indeed, I think that postulating such lexical ambiguities is a very doubtful practice. Of course, with respect to disjunctions it can be claimed that all that is needed is symmetric disjunction. Indeed, this disjunction seems to have the right cancellation effects. But I don't see how it's interpretation rule can be motivated in an independent way. Symmetric disjunction seems natural in case the disjunction is *exclusively* used. In that case it can indeed be assumed that what is asserted by each disjunct is done on the assumption that the other disjunct is false. But we would face two well-known problems if we would assume that disjunctions should be interpreted exclusively: it is wrongly predicted that *A or B* is equivalent to *~A or ~B*,[221] and if *or* denotes a binary connective, an assertion of the form *A, or B, or C* will still be true if all three disjuncts are true. With respect to the other cases, it seems more reasonable to assume that presuppositions can be cancelled for reasons of informativity.

4.5 Cancellation by informativity

According to any truth-conditional account and to the pure satisfaction account towards presuppositions, a sentence always gives rise to one presupposition, and whether it puts constraints on contexts depends on whether or not the presupposition is trivial. Although the sentence *It will stop raining* can only appropriately be uttered in contexts where it is assumed that it was raining, the conditional

(3c) If it was raining, it will stop raining.

does not seem to constrain contexts. According to the truth-tables of § 4.2 and the satisfaction approach this is because (3c) gives rise to the presupposition *If it was raining, it was raining*. Because this denotes a non-informative proposition, no presuppositional constraint is put on the context. However, we have seen that there are other examples that intuitively do not presuppose anything, although Peters' truth tables, and the original satisfaction account would predict a presuppositional reading. We have seen that stipulating ambiguities of *not, if* and *or* technically can solve the problem, but most authors find such a move unsatisfactory. Let's consider (6) again

(6) If I realise later that I have not told the truth, I will confess it to everyone.

The pure satisfaction account predicts that (6) presupposes that the speaker has not told the truth. But of course, the sentence doesn't have this presuppositional reading. Why is that? The reasoning is due to Stalnaker (1974): If a speaker explicitly supposes *A* by his use of *A* as the antecedent of an indicative conditional, he indicates that he is not assuming that *A* is already presupposed. Why else supposing it? For (6) this means that the antecedent of the conditional is not already presupposed; the speaker does not yet presuppose that he realises that he hasn't told the truth. But now suppose that the calculated presupposition of (6) would really be presupposed by the speaker. That would mean that the speaker would

---

[221] The sentence *John either wears a coat or a hat* seems not to be truth-conditionally equivalent with *John either wears no coat, or no hat.*

presuppose that he has not told the truth. Assuming that when somebody presupposes something, it can no longer be an issue for him whether it is true or not, and that somebody realises that A iff A holds and he also believes that A, the presupposition calculated by (6) is in contradiction with the implicature induced by the making of the supposition that what is explicitly supposed is not already presupposed. Thus, the presupposition is cancelled.

To account for (6), three things are crucial: (i) the sentence by which an assertion is made is part of the context needed to determine what is presupposed by the assertion of this sentence, (ii) the assumption that presupposition is a propositional attitude, and (iii) the assumption that some potential presuppositions can be cancelled. With respect to the first point, it should be uncontroversial that one of the facts available for the participants of the conversation to interpret an utterance is the fact that the utterance is made with the particular words used. The second assumption we have made all along; what is presupposed is the presumed common knowledge or common belief of the participants of the conversation.[222] The third assumption is controversial, but I think perfectly acceptable for at least some presupposition triggers.

Some presuppositions can be cancelled, but certainly not all of them. It is commonly assumed that the presupposition associated with the word *even* is not entailed by the sentence. Suppose that the calculated presupposition of *even* is cancelled. The result would be that the word *even* does not only have no effect on what is asserted, but has also no effect on what is presupposed. In that case it would be unclear why the speaker used the word in the first place. For this reason Karttunen & Peters (1979) claimed that these presuppositions, or conventional implicatures as they called them, cannot be cancelled for reasons of informativity. This seems to hold in general, presuppositions that are not entailed cannot be cancelled for reasons of informativity. It is also hard to imagine that the presupposition of a referentially used description like *the man* can be cancelled by informativity; it would not be clear what would be expressed by the sentential clause in which the description occurs. I will assume that only those triggers give rise to presuppositions that can be cancelled that put only constraints on the information about the subject matter of conversation, and whose use can be explained even if the associated presupposition is cancelled. Typical examples are attributively used descriptions, factive verbs, and maybe aspectual verbs like *stop*.[223] Even those kind of presuppositions that can be cancelled, can only be cancelled in exceptional cases. The speaker must make it clear to the other participants that he is not making the relevant presupposition. Is it possible to give general rules by which we can calculate when a potential presupposition is cancelled and when not? And if we can give such rules, can these rules be independently motivated by conversational principles? Gazdar (1979), Soames (1979, 1982) and Van der Sandt (1982, 1988) have proposed that both questions can be affirmatively answered. The proposal comes down to this: *If* each sentential clause would be neither inconsistent nor trivially true with respect to its context of interpretation as given by the CCP rules when the initial context for the whole sentence is accommodated with the calculated presupposition(s) of the sentence, *then* the sentence has the presuppositional reading, otherwise not. The condition for acceptable accommodation is basically Stalnaker's (1978) first assertion condition: A proposition asserted is always true in some but not all of the possible worlds in the context set. Note that according to this proposal, we have to be able to determine what is potentially presupposed by a sentence independently of what is asserted.

---

[222] Besides Stalnaker (1973, 1974) there are not so many authors who take the assumption serious that presuppositions are propositional attitudes. Gazdar (1979) and Blok (1993) are two exceptions.

[223] This does not mean, of course, that these expression cannot contain indexical or referential expressions. For instance, the description *The king of this country* can be used attributively.

4.6 A combined two-dimensional approach[224]

According to proponents of the pure satisfaction account, a presupposition associated with a part of a sentence need not be a presupposition of the whole sentence, because the presupposition of this part is already entailed by another part of the sentence. Thus presuppositions can be *cancelled*[225] *by entailment*. According to proponents of the pure cancellation account, a presupposition associated with a part of a sentence need not be a presupposition of the whole sentence, because if the presupposition of the part would be also a presupposition of the whole, not every assertive part of the sentence would be *consistent* and *informative* with respect to what is already presupposed. Presuppositions are *cancelled by informativity*. If we assume that presuppositions can be cancelled by informativity, the question arises whether we still need something like *satisfaction*. Gazdar (1979), Soames (1979) and Van der Sandt (1982) do not make use of the latter notion. None of these authors filtered presuppositions out by entailment; presuppositions are only cancelled for reasons of informativity. But that proposal was wrong as first seen by Soames (1982). If ∂ is the presupposition operator (Beaver, 1992), they predict wrongly that sentences of the form $A \rightarrow (\partial P \wedge B)$, where $A \models P$, but $P \not\models A$ presuppose P. Thus, *If John has sons, then Mary will not like his children* is wrongly predicted to unambiguously presuppose that John has children.[226] Soames (1982) observed that the theories that account for cancellation by entailment, and theories that account for cancellation for reasons of informativity were largely complementary, and concluded that any successful theory has to use both ways of cancelling presuppositions. As discussed by Soames (1982), there are at least two ways of combining these two approaches. The easiest, and most Gazdarian combination can be characterised by the slogan *if not entailed, then accommodate, if acceptable*.[227] Let us now state a theory that uses both cancellation by entailment and by informativity, but only consider information about the subject matter of discourse. In this section I will ignore referential expressions, and assume that what is asserted by a sentence is a proposition, a function from worlds to truth-values, that can be determined in a (reference-) context-independent way.

In contrast to the theory of Karttunen & Peters (1979), in the first statement of my account, the representation of an utterance contains both assertive and presuppositional material. I follow Beaver (1992) in assuming that ∂, the presuppositional connective, has only atomic propositions in its domain.[228] I assume that ∂ is really a collection of two unary operators, \$, and §. If \$A is meant by ∂A, then A also counts truth-functionally, if §A is meant, then not.[229] For instance, a sentence like *If John regrets that A, then B* is represented as (\$A $\wedge$

---

[224] It's sometimes (e.g. Krahmer, 1995) claimed that two-dimensional theories can be replaced by four-valued partial logics. This is misleading. Yes, the claim is true for the theory of conventional implicatures as stated in Karttunen & Peters (1979), but only because they assume that presuppositions cannot be cancelled and because of the specific presuppositions they calculate for complex sentences. If cancellation is allowed, or if we assume the two-dimensional account of Karttunen (1973), no many-valued logic will be equivalent. Moreover, there is no principled reason to assume why the presuppositional dimension, for instance, should be two-valued, and not three-valued instead. In sum, the theory of Karttunen & Peters (1979) is not the only possible two-dimensional presupposition theory.

[225] Or made trivial.

[226] According to Van der Sandt (1989, 1992), the sentence still might have this presupposition.

[227] This account is obviously closely related with Van der Sandt's (1992) theory, which can be characterised by the slogan *if not bound, then accommodate, if acceptable*. By taking salience serious, we can say that binding is really entailment.

[228] I will assume that presupposition requirements are written into the lexical entries of particular words. If presuppositions are embedded in other presuppositions, the most deeply embedded presuppositions will always stand in front. Thus, the sentence *John knows that the king of France is bald* would be represented by something like '\$E(ιxKF(x)) ∧ \$bald(ιxKF(x)) ∧ Know(john, bald(ιxKF(x)))', if the description is attributively used, and where 'E' is the existence predicate.

[229] Very roughly, the presuppositions called *conventional implicatures* by Karttunen & Peters (1979) I represent by §, whereas the other presupposition triggers are represented by \$.

Regret(A)) → B. Also the presuppositions of other factives, aspectual verbs like *stop*, and attributively used definite descriptions (or diagonalised proper names) are represented by means of $; thus, if the description *the king of France* is attributively used, the sentence *The king of France is bald* presupposes that France has a king, but is simply false at the actual world. On the other hand, the presuppositions associated with, for instance, *even* and *also* are represented by means of §. If it is the case that Bill likes Mary, but not that Bill is the least likely person to like Mary, the sentence *Even Bill likes Mary* is considered to be true, although the presupposition does not hold. The account is two-dimensional in that it gives, just like Karttunen (1973), Karttunen & Peters (1979), Gazdar (1979), and Soames (1982), separate recursive definitions of what is asserted and of what is presupposed by an utterance. That is, it depends partially on what is asserted which potential presuppositions actually get accommodated.

Let us assume for simplicity that what is said by an utterance can be determined in a completely context-independent way. Then, given a model $\langle W, \models \rangle$, we can determine the static interpretation function, $[[.]]^{st}$, for utterances in the standard Boolean way, plus the following definitions:

$[[\$A]]^{st} = [[A]]^{st}$
$[[\S A]]^{st} = W$

Let S be a sentence, the question is which potential presuppositions of S have to be accommodated to a particular context. For simple sentences the question is easily answered: If S is a simple sentence, then the potential presuppositions of S that have to be accommodated are those potential presuppositions of the parts of S not entailed by the context such that assuming these presuppositions does make the assertion of S neither trivial nor inconsistent with respect to what is already assumed. For the determination of the to be accommodated presuppositions of complex sentences, things are slightly more complicated. By the determination of potential presuppositions of, for instance, conjunctions, the left to right asymmetry becomes relevant. The presuppositions of $A \wedge B$ (relative to S) will be the presuppositions of A (relative to S) that are not entailed by the context, plus those presuppositions of B (relative to S) that are not entailed by the context updated with A. This is almost the same as Karttunen (1973), and Stalnaker (1973, 1974) argued for. The pragmatic explanation for the projection behaviour of presuppositions in complex sentences can be used, and was even originally intended, to give an independent motivation for the Karttunen (1973) rules for calculating the (potential) presuppositions of complex sentences.

We have seen that context K is used to determine which of the potential presuppositions associated with a sentence used to make an utterance have to be accommodated. A potential-presupposition P of A has no influence on its context of interpretation K, if K already contains the information that P, that is if $\cap K \subseteq P$.

For the interpretation of A in K disregarding pragmatic constraints, the following function can be defined:

$\{A\}(K) = \{[[A]]^{st} \cap s \mid s \in K\}$

According to the cancellation account, we first determine the set of potential presuppositions of a sentence, and then determine which of those potential presuppositions is also an actual presupposition, and must be accommodated. As explained above, a potential presupposition of a sentence is also an actual presupposition of a sentence, if every assertive part of the sentence would be *consistent* and *informative* with respect to this potential presupposition plus what is asserted by the 'earlier' parts of the sentence. To formalise this, we will determine for each sentence a set of acceptable initial contexts of

interpretation. If what is expressed by a potential presupposition is an element of this set, the potential presupposition will not be cancelled. Following some suggestions made by Stalnaker (1973, 1974, 1975), and Van der Sandt (1982, 1988), the set of acceptable contexts with respect to informativity for [[A]], $\mathbf{A}(A)$, is construction dependent recursively defined as follows:[230]

$\mathbf{A}(A) = \{s \subseteq W \mid (\llbracket A \rrbracket^{st} \cap s) \neq \emptyset \ \& \ (\llbracket A \rrbracket^{st} \cap s) \neq s\}$, if A is atomic
$\mathbf{A}(\sim A) = \mathbf{A}(A)$
$\mathbf{A}(A \wedge B) = \{s \subseteq W \mid s \in \mathbf{A}(A) \ \& \ (\llbracket A \rrbracket^{st} \cap s) \in \mathbf{A}(B)\}$
$\mathbf{A}(A \rightarrow B) = \mathbf{A}(A \wedge B)$, if $\mathbf{A}(A \wedge B) \neq \emptyset$
$= \mathbf{A}(B)$ otherwise[231]
$\mathbf{A}(A \vee B) = \{s \subseteq W \mid (\llbracket \sim B \wedge A \rrbracket^{st} \cap s) \neq \emptyset \ \& \ (\llbracket \sim A \wedge B \rrbracket^{st} \cap s) \neq \emptyset\}$
$\mathbf{A}(\partial A) = \wp(W)$

The assumption that the speaker has a context in mind for the utterance of A that is an element of $\mathbf{A}(A)$ can be thought of as the prior presupposition that the speaker will not claim something that is already accepted to be trivially true or false.

Partly on the basis of this prior assumption we can now determine which potential presuppositions have to be accommodated, the *actual* presuppositions of a sentence. Roughly speaking, for each potential presupposition of a clause we first check whether the presupposition is not cancelled because it is entailed by the initial context incremented with the earlier parts of the sentence, and then check whether the presupposition is not cancelled for reasons of informativity by means of the above defined acceptability relation. I will define $f_p(A, K, S)$, the set of actual presuppositions of A in context K relative to utterance S. $f_p(A, K, S)$ will be a set of *logical forms*, not a set of propositions. It was commonly assumed in the two-dimensional presupposition theories that it didn't matter whether what we called $f_p(A, K, S)$ is a set of logical forms or a set of propositions.

$f_p(A, K, S) = \emptyset$, if A is atomic
$f_p(\sim A, K, S) = f_p(A, K, S)$
$f_p(A \wedge B, K, S) = f_p(A, K, S) \cup f_p(B, \{A\}(K), S)$[232]
$f_p(A \rightarrow B, K, S) = f_p(A, K, S) \cup f_p(B, \{A\}(K), S)$[233]

---

[230] In Stalnaker (1975), **A** denotes an appropriateness relation between sentences and contexts. I will say that **A** denotes a function from sentences to sets of contexts, but the two are obviously equivalent.

[231] Note that for a sentence like *If John knows the war is over, the war is over* that is of the form '$A \rightarrow B$', $\mathbf{A}(A \wedge B) = \emptyset$. It has been suggested by Landman (1986) that such a sentence can only be appropriately used to make clear what follows from the antecedent. I agree, but think that something else is special about the sentence, too. I believe that such a sentence can be uttered appropriately only if the antecedent has been uttered and accepted before. The conditional is still informative, because not every participant of the conversation knows what proposition is expressed by the sentence. The conditional expresses meta-linguistic information about the words used in an earlier sentence, repeated by the antecedent.

[232] Note that we now predict, following Karttunen (1973), that a sentence uttered in an empty context that is represented by the formula '$A \wedge (\partial P \wedge B)$' (or '$A \rightarrow (\partial P \wedge B)$'), presupposes that *P*, if *A* does not entail *P*. According to Karttunen (1974), however, the sentence should have the conditional presupposition *If A, then P*. Only for convenience I have chosen the former account. In fact, I prefer the latter, and we can easily account for the latter intuition, too, in our two-dimensional framework. Although the issue how to determine the presuppositions of these sentences has been given a lot of attention recently, see especially Beaver (1993) and Geurts (1995), I prefer to ignore it here.

[233] Van der Sandt (1992) argued that this way of calculating the presupposition of a conditional cannot be correct. It would predict, he thinks, that a sentence like *If John has twelve grandchildren, his child will be happy* does not give rise to a presuppositional reading. This is true if the expression *John's child* gives rise to the *existential* presupposition that John has a child. But this assumption is simplistic. If it is assumed

$f_p(A \vee B, K, S) = \{\sim(A \wedge B)\} \cup f_p(A, \{\sim B\}(K), S) \cup f_p(B, \{\sim A\}(K), S)$,
if $[[\sim(A \wedge B)]]^{st} \in A(S)$
$= f_p(A, K, S) \cup f_p(B, K, S)$ otherwise[234]

$f_p(\partial A, K, S) = \{A\}$, if $\cap K \not\subseteq [[A]]^s$ & $\exists s \in \{A \wedge \text{Presup}(A)\}(K): s \in A(S)$[235]
$= \varnothing$ otherwise.

Now we can determine $\Uparrow(A, K, S)$, the *presupposition* of A relative to K and S:

$\Uparrow(A, K, S) = \cap\{[[C]]^{st} \mid C \in f_p(A, K, S)\}$, if $f_p(A, K, S) \neq \varnothing$
$= W$, if $f_p(A, K, S) = \varnothing$

A context set K really consists of a set of contexts. Following Beaver (1992), the interpretation of a sentence A in a context K is the distributive update of A with respect to the acceptable contexts for A in K:

$[A](K) = \{[[A]]^{st} \cap s \mid s \in K \,\&\, s \in A(A) \,\&\, s \subseteq \Uparrow(A, K, A))\}$

So, accommodation of presupposition (= filtering of context set) is done before the context is updated with what is asserted. Beaver's (1992) idea was that semantic and pragmatic information are analysed at a different level. Semantic information is analysed at the local level, while pragmatic information is analysed at the global one. Note that [.] is a total function in $[(L \times \wp(\wp(W))) \rightarrow \wp(\wp(W))]$. For what kind of expression A do we look at [A](K)? Beaver (1992) assumed that for each clause [.] should be defined. In two dimensional theories, on the other hand, it is assumed that such an A is never a constituent of a sentence, because it is that which is asserted in one speech act. I will assume that such an A is minimally of the sentential level, but it might also be more coarse-grained.

In this two-dimensional account we can define different kinds of entailment relations:

$A \models B$	iff	$[[A]]^{st} \subseteq [[B]]^{st}$	semantic entailment
$A >> B$	iff	$\cap A(A) \subseteq [[B]]^{st}$	clausal implicature[236]
$A \text{->}_K B$	iff	$\Uparrow(A, K, A) \subseteq [[B]]^{st}$	presupposition
$A \Rightarrow_K B$	iff	$\cap[A](K) \subseteq [[B]]^{st}$	reasonable inference

Let us illustrate our machinery by considering two examples. First, consider the classical *The king of France is not bald* in empty context K, thus where $K = \wp(W)$. The sentence

---

that it gives rise to the presupposition that there is a unique most salient child in all possibilities of the context, it can no longer be assumed that the antecedent entails the presupposition of the consequent.

[234] So, a disjunction is either exclusively used, in which case both disjuncts are asserted on the assumption that the other disjunct is false, or it is not, in which case the context of interpretation for each disjunct is the whole context. As a result, *Either it storms, or the king of Buranda will come on time* presupposes that there is a king of Buranda, but the potential presuppositions of *Either John stopped smoking, or he just started doing so* are cancelled by entailment.

[235] Karttunen (1973) only filtered presuppositions out by entailment, he did not use acceptability conditions. But that was wrong, he predicts wrongly that sentences of the form $A \rightarrow (\partial P \wedge B)$, where $P \models A$, but $A \not\models P$ presuppose P. Thus "If John has children, then *his sons* have red hair" is wrongly predicted to presuppose that John has sons. I now follow Gazdar (1979), Soames (1979) and V.d.Sandt (1982, 1992) in saying that this cannot be a presupposition because assuming it would make the antecedent uninformative in its context of interpretation.

[236] After Gazdar (1979).

will be represented in such a way that the description has smaller scope than the negation: ~($Exist(ιxKFx) ∧ Bald(ιxKFx)). We want to determine f_p(~($Exist(ιxKFx) ∧ Bald(ιxKFx)), K, ~($Exist(ιxKFx) ∧ Bald(ιxKFx))), which is the same as f_p($Exist(ιxKFx) ∧ Bald(ιxKFx), K, ~($Exist(ιxKFx) ∧ Bald(ιxKFx))). Because the second conjunct does not contain a potential presupposition, the latter will be the same as f_p($Exist(ιxKFx)), K, ~($Exist(ιxKFx) ∧ Bald(ιxKFx))). Because K does not entail, nor is inconsistent with the assumption that there is a king of France, and because the assumption that there is a king of France does not rule out nor entail that this king is bald, it is predicted that the sentence actually presupposes that there is a king of France, as desired. As a second example we will explain why (8),

(8) If I discover that my wife has been playing around, I will be upset. (Stalnaker, 1974)

does not presuppose that my wife has been playing around. Sentence (8) will be represented by the following logical form (neglecting the presupposition associated with *my wife* and assuming that the conditional should be represented by the material implication): '$(my wife has been playing around) ∧ Regret(I, my wife has been playing around)) → I will be upset'. We want to be able to cancel the potential presupposition that my wife has been playing around. We have to determine whether $\hat{\Pi}$($(my wife has been playing around) ∧ Discover(I, my wife has been playing around)) → I will be upset, K, $(my wife has been playing around) ∧ Discover(I, my wife has been playing around)) → I will be upset) ⊆ [[my wife has been playing around]]st. This will be the case iff the sentence 'my wife has been playing around' will be an element of f_p($(my wife has been playing around), K, $(my wife has been playing around) ∧ Discover(I, my wife has been playing around)) → I will be upset). In order for that to be true, it has to be the case that [[my wife has been playing around ∧ Presup(I, my wife has been playing around)]]st ∈ A($(my wife has been playing around) ∧ Discover(I, my wife has been playing around)) → I will be upset). But this won't be true if it is assumed that what one presupposes already, one cannot discover later. As a result, the presupposition is cancelled for reasons of informativity, as desired.

Note that according to the above system, B can be reasonably inferred from A given K, if B is semantically entailed by A, B is a clausal implicature of A, or if B is a presupposition of A given K. Observe also that we can reasonably infer from the assertion that A is the case, that both A and not-A are consistent with what is presupposed, if A is not valid.[237] The same holds for the assertion that ~A is the case. Similarly, if it is asserted that A or B is the case, we can infer from the utterance made that both A ∧ ~B, and ~A ∧ B are possible as far as what the speaker presupposes, if A and B are logically independent of each other. If it is assumed that the description *the king of France* is attributively used, we predict that with an utterance of a sentence like *The king of France is bald* it is both presupposed and asserted that there is a king of France. On the other hand, an utterance of *The king of France (does not) exist(s)* does not give rise to such a presupposition, because that would make the assertion trivially true (false).

If we don't think of presuppositions as definedness conditions, we can assume that scalar implicatures are treated as presuppositions, and, moreover, that their presupposition can be

---

[237] If A is valid, the speaker probably wants to assert that the (horizontal) proposition expressed by A is the necessary proposition. In that case, it is presupposed that the speaker does not know this meta-linguistic information.

cancelled. Thus, *A or B* and *Some S are P* presuppose respectively that not A and B, and that not all S are P. Note that by the acceptability rules and presupposition together, the exclusive reading of *A or B*, either A and ~B holds, or ~A and B, is reasonably inferred.[238] The presuppositions triggered by disjunctions and indefinites are cancelled in sentences like *A or B or both* and *Some S are P, if not all.* Thus, the addition of the third conjunct in *A or B or both* adds nothing to the assertive content of the disjunction, but says only something about the context, about what is presupposed by the speaker.

In this section we have assumed that with negative sentences we always make an assertion, and that the potential presuppositions associated with sentences like *the king of France does not exist, The king of France is not bald, because there is no king of France*, and *John doesn't regret failing, because in fact he passed* are cancelled for reasons of informativity. This was possible because we assumed that descriptions can be used attributively, and that the potential presuppositions associated with descriptions and factive verbs are also entailed. Indeed, I believe that at least for some uses of those sentences this is the right analysis, but normally something different seems to be going on. This comes out clearly if we consider the following example that is very similar to the ones above:

(9a) A: It is possible that the church is right.
(9b) B: It is not *possible*, but *necessary* that the church is right.

Although it might be reasonably inferred from (9a) that it is not necessary that the church is right, it is commonly assumed that this is not semantically entailed by the sentence. But if it is also assumed that each conjunct of (9b) is assertively used, what is asserted by the second conjunct will be in contradiction with what is asserted by the first conjunct, which is absurd. It seems more reasonable to follow Van der Sandt (1991) and assume that the first conjunct of (9b) is not assertively used, but used, instead, to reject a previous utterance.[239] Indeed, the crucial observation, I think, is that sentences like (5) cannot be uttered very naturally in the beginning of a conversation.[240] According to Van der Sandt (191), we normally anaphorically refer back to a previous utterance with such a sentence, and reject all information reasonably implied by this previous utterance that was not already presupposed.

If we assume that negation can be used in two ways; to make an assertion with a negative sentence, and to make a denial, we have to assume that the negation sign in natural language is ambiguous. According to Karttunen & Peters (1979) and Seuren (1985), for instance, the negation sign is *semantically* ambiguous. It is proposed that one negation is presupposition preserving, while the other is not. But this proposal gives rise to two problems: First, if one assumes that negation is semantically ambiguous, one would expect there to be a language that has two distinct morphemes for the two underlying negation operators, which does not seem to be the case (cf. Gazdar, 1979, pp. 65-66). Second, even if the negation in (9b) is used in a non-presupposition preserving way, it still cannot be explained why the first *and second* conjunct of (9b) can be appropriately used in the context of (9a). If (9a) conveys the information that it is not necessary that the church is right, what is expressed by the second conjunct of (9b) would still be inconsistent with its context of interpretation, if the information conveyed by (9a) is not retracted from the context.

A more attractive solution is proposed by Van der Sandt (1991). He proposes that the negation sign is not *semantically*, but *pragmatically* ambiguous. According to Stalnaker

---

[238] The same effect is reached by Gazdar (1979), but, as noted by Groenendijk & Stokhof (1984, pp. 424-25), on the basis of a much too strong scalar implicature.

[239] See also Stalnaker (1978).

[240] See Seuren (1985), and Van der Sandt (1991) for much more observations concerning the use of negation to reject a previous utterance.

(1970b), a sentence can be pragmatically ambiguous, if some rule involved in the interpretation of that sentence can be applied either to the context, or to the index. We have already seen that sentences can be pragmatically ambiguous because the description occurring in the sentence can be used referentially and attributively. Another example mentioned by Stalnaker are sentences of the form *If A, then B*; they can be used to assert a conditional proposition, and to assert *B* on the condition that *A* is the case, otherwise nothing is asserted. In a similar way, Van der Sandt proposes that also sentences of the form *It is not the case that A* are pragmatically ambiguous. They can be used to assert a negated sentence, or to reject the new information reasonably implied by a previous utterance.[241] In case the negation is used as a *denial*, to reject a previous statement, the clause in which the negation occurs is not used to introduce new information about the subject matter of conversation, but to *retract* some information explicitly introduced earlier. If it is assumed that with the first conjunct of (9b) the speaker intends to retract the information introduced by (9a), it can be explained why he can use the second conjunct of (9b) appropriately; after the retraction, the information the second conjunct expresses will be consistent with the context in which it is interpreted.[242]

4.7 Attitudes

In chapters 1 and 3 we have extensively discussed belief attributions. We focused mainly on expressions like demonstrative and anaphoric pronouns that are *referentially* used in embedded sentences of belief attributions. We made a lot of use of *double indexing*; referring expressions used in the embedded sentence of a belief attribution were interpreted *in situ*, but by making use of counterpart functions, we could still refer to objects introduced in the main context, because the reference-context is always the reference-context of the main context. Even to account for anaphoric relations across belief contexts, we still could assume that the reference-context is always the reference-context of the main context, because we proposed that indefinites in belief contexts can introduce belief objects. Just as in the last section, we will here ignore referential expressions, and thus ignore the technique of double indexing. A context will simply be represented by a set of indices, and we will almost only consider triggers that give rise to presuppositions that don't have to be considered as definedness conditions.

What happens if the fragment of the language also contained some propositional attitude verbs? Given that DOX(a,w) is the set of worlds compatible with everything what *a* believes in *w*, the static interpretation rule for belief sentences, for instance, can easily be given:

$$[[\text{Bel}(a, A)]]^{st} \quad = \quad \{w \in W \mid \text{DOX}(a,w) \subseteq [[A]]^{st}\}$$

What about presuppositions? Until now we have assumed with Gazdar (1979) that the set of presuppositions of a complex sentence is a subset of the set of presuppositions of its parts. If we generalise that idea to attitudes, if somebody utters *John believes that it stopped raining* out of context, he presupposes that it is in fact raining, because that's what is presupposed by the embedded sentence. But as first noticed by Van der Sandt (1982, 1988), things are not that easy. Although it's not implausible to assume that belief sentences normally presuppose the presuppositions of their embedded clauses, this is certainly not always the case. A sentence like

(10) If Louise believes that it was raining, then she also believes that it *stopped* doing so.

---

[241] See also Horn (1985) for a similar claim, although he accounts for denials in a purely metalinguistic way.

[242] See Van der Sandt (1991) for a formal account of denials in a framework closely related to the one presented here.

does not presuppose that it was raining, although this would be predicted by the above Gazdarian reasoning. It seems that in a sentence like A @ Bel(a, B), where @ is either a conjunction or an implication, a presupposition of B does not influence the main context nor the belief context, if it is entailed by either A, or by the belief context determined by A. The belief context determined by A is that what is presupposed to be compatible with what the agent believes after A is a accepted. The latter context is called the *derived belief context* by Stalnaker (1988) and determined as follows (for the agent called 'Phoebe'):

> for each possible situation in the basic context, Phoebe will be in a definite belief state which is itself defined by a set of possible situations - the ones compatible with what Phoebe believes in that possible situation. The union of all the possible belief states will be the set of all possible situations that might, for all the *speaker* presupposes, be compatible with Phoebe's beliefs. (Stalnaker, 1988, p. 146)

Because in my account a context, K, is a set of sets of possible worlds, I only follow the proposal of Stalnaker at one level higher. Let us say that if *a* is an agent, and *w* a world, there is a doxastic accessibility relation DOX(a,w). In terms of such an accessibility relation for worlds, the accessibility relation for a set of worlds can be determined: DOX(a,s) = $\{DOX(a,w) | w \in s\}$. Thus, if we say that a set of worlds is a situation, DOX(a,s) consists, just like K, of a set of situations. The belief context DOX(a,K) for *a* derived from the main context, K, is determined as follows: $DOX(a,K) = \cup\{DOX(a,s) | s \in K\} = \{DOX(a,w) | \exists s \in K: w \in s\}$. Note that also DOX(a,K) consists of a set of situations. Now we can determine $f_p(Att(a, A), K, S)$ for any attitude verb 'Att':

$$f_p(Att(a, A), K, S) = \{C \in f_p(A, DOX(a,K), S) | \cap K \not\subseteq [[C]]^{st}\}$$

In this way the potential presuppositions of the embedded sentence are also presuppositions of the whole sentence, if they are neither entailed by DOX(a,K), nor by K. Thus, the potential presuppositions triggered by *stopped* in the consequents of the following two sentences are both cancelled by entailment:

(11) If Fred used to beat his wife, then Bill hopes that Fred will *stop* beating her.
(12) If Bill believes that Fred used to beat his wife, then he hopes that Fred will *stop* beating her.

Consider (12), for instance, of the form 'Bel(b, A) $\rightarrow$ Hope(b, \$A $\wedge$ B)' and uttered in empty context K. The derived belief context for Bill of K updated with Bel(b, A), K', will be {A}(DOX(a, K)), and thus contains the information that A. Because $\cap DOX(b, K') \subseteq [[A]]^{st}$, the presupposition associated with the consequent of (12) will be satisfied in its local context of interpretation, and thus will not be accommodated to the main context.

For the Gazdarian approach to work, we also have to define the acceptability conditions for belief sentences. According to Stalnaker (1988), a belief sentence is acceptable if and only if the derived belief context is an acceptable context for the embedded clause:

$$\mathbf{A}(Bel(a, A)) = \{s \subseteq W | \cup\{DOX(a,w): w \in s\} \in \mathbf{A}(A)\}$$

According to the Gazdarian-like analysis presented above, presuppositions of embedded sentences of attitude attributions are accommodated to the main context, if acceptable, if they are not entailed by either the main or the derived context. However, there is a serious problem with this proposal, which shows in sentences like

(13) Louise believes that the king of France is bald, although there is no king of France.

According to our above rule the presupposition of the embedded sentence is cancelled, but then the question arises how the embedded sentence can be interpreted. Of course, Gazdar (1979) has no real problem with a sentence like (13), if it is assumed that the description has smaller scope than the belief predicate, and if the description does not only give rise to the presupposition that there is a king of France, but that this is also semantically entailed. However, the assumption that the rule determining the denotation of an attributively used description is part of the proposition expressed by the embedded sentence of a belief attribution, together with the assumption that all definite descriptions give rise to presuppositions that the description *actually* has a unique denotation, gives rise to two problems. First, as Van der Sandt (1992) observed, there is a *binding problem*. Consider the assertion of (14) in a situation compatible with there being a pope:

(14) John believes that *the pope* comes from Holland.

On the basis of the above given rule for determining the presuppositions of belief attributions it is predicted that the sentence actually presupposes that there is a pope. But if we also make the Gazdarian assumptions that what is presupposed by a definite description is also entailed and that definite descriptions always get semantically the smallest scope possible, it follows that for a sentence like (14), the utterance has in a sense both a *de re* and a *de dicto* reading *at the same time*. But now we have a close variant of Kaplan's problem of *the shortest spy*. If *a* is the actual pope and John is acquainted with *a*, but still thinks of somebody else as the pope, it is predicted that John thinks of this other person that he is the pope who comes from Holland. Moreover, nothing is predicted about what John thinks about *a*, the actual pope. But both results are counter-intuitive. If the utterance presupposes that there is a pope, we should predict that it is this pope that John has beliefs about.

Of course, this problem would disappear if the Gazdarian assumption that definite descriptions always get the smallest possible scope is given up. Indeed, I believe that for all presupposition triggers we should first determine the scope of the trigger relative to the attitude verbs, if they are scope sensitive at all, and only then determine the presupposition that the trigger gives rise to. But then we should also give a different definition of f_p(Att(a, A), K, S), because we should no longer assume that a presupposition of the embedded clause, A, that has nothing to do with the reference-context can put constraints on the main context. Instead, we should say that such presuppositions only constrain the context of interpretation for the embedded context: that what is presupposed to be believed by the agent. As a result, a presupposition trigger used in the embedded sentence of a belief attribution gives rise to, as a rule, the constraint that the agent believes in the truth of the associated presupposition. This is in accordance with the proposals of Karttunen (1974), Karttunen & Peters (1979), Stalnaker (1988), and Heim (1992). This intuition can be implemented in a two-dimensional framework by the following definition:

$$f_p(\text{Att}(a, A), K, S) = \{\text{Bel}(a, C) \mid C \in f_p(A, \text{DOX}(a,K), \partial S) \,\&\, \text{Bel}(a, C) \in \mathbf{A}(S)\}^{243}$$

The above definition does not only help to solve our above binding problem, it also escapes another problem Gazdarians have to face (first observed by Van der Sandt (1982)). We have seen that in two-dimensional accounts that crucially make use of presupposition cancellation it is normally assumed that the presupposition triggered by factive verbs is also entailed. Thus, *John knows that A* is false when A is not the case, although *Sam doesn't know that A* can still be true in this case. For a sentence like

(15) Mary believes that John knows that his vacuum cleaner is broken.

---

[243] Note that by the use of ∂S in f_p(A, DOX(a,K), ∂S), all-potential presuppositions of A not entailed by DOX(a,K) will be elements of f_p(A, DOX(a,K), ∂S). That is, nothing is cancelled for reasons of informativity.

uttered out of context, it is predicted by the Gazdarian definition of $f_p(Att(a, A), K, S)$ that (15a) is asserted, and (15b) presupposed:[244]

(15a) Mary believes that John has a vacuum cleaner that is broken, and that John knows this
(15b) John has a vacuum cleaner that is broken.

This is already strange enough, but now consider:

(16) Mary hopes that John knows that his vacuum cleaner is broken.

It is predicted that out of context with this sentence (16a) is asserted and (16b) presupposed:

(16a) Mary hopes that John has a vacuum cleaner that is broken, and that John knows this.
(16b) John has a vacuum cleaner that is broken.

Thus, what is asserted by (16) can be represented by Hope(m, $A \wedge B$). But given that the presupposition is only (16b), this is problematic; with (16) we normally don't want to say that Mary hopes that John has a broken vacuum cleaner. We want to say that Mary hopes that John knows that his vacuum cleaner is broken, *given* that Mary believes that John's vacuum cleaner is broken.[245] It is not clear how this intuition can be accounted for if it is predicted that (16b) is the presupposition of (16). On the other hand, if what is presupposed by an attitude attribution is that the agent must already believe that the presupposition is true, the intuition is accounted for in a straightforward way, if what one desires depends on what one believes.[246]

Thus, we can claim that presupposition triggers used in embedded clauses of attitude attributions give rise to the presupposition that the agent *believes* the associated presupposition of the embedded clause. Note that once we have accepted this rule, we have made an important step in the direction from Gazdar to Karttunen & Peters. According to Gazdar, the set of presuppositions of a complex sentence is a subset of the set of presuppositions of its parts. On the other hand, according to Karttunen & Peters the presupposition(s) of a simple clause can be transformed into a more complex presupposition of a more complex sentence of which this simple clause is a part. To account for the presuppositions associated with attitude attributions, it seems we have to follow the latter approach.

Still, if we propose the last definition of $f_p(Att(a, A), K, S)$, we seem to have a problem. How can we account for the fact that a sentence like

(17) John believes that it stopped raining

still invites the inference that in fact it was raining?[247] I believe this has something to do with the reason *why* we make attitude attributions.

---

[244] Even if factives are treated as scope bearing operators, it seems that the factive should have small scope with respect to the belief operator.

[245] See also Stalnaker (1984) and Heim (1992) for similar observations. Stalnaker's example is very convincing: if I say *I want to know who committed the murder*, I don't want the murder to be committed, but only want to know who committed the murder *given* that the murder is committed.

[246] See chapter 6 for more discussion.

[247] Zeevat (1992) claims that a presupposition trigger in the scope of an attitude sentence gives rise to the presupposition triggered by the embedded sentence itself, *and independently* that this presupposition is believed by the agent. Unfortunately, I don't see how to account for this without getting into binding problems.

It is sometimes assumed that every assertion is an answer to an implicit or explicit question. Attitude attributions shouldn't be exceptions. For instance, the topic of a discourse can be who committed the murder. The assertion *John committed the murder* is an appropriate claim to make, but so is *Mary believes that John committed the murder*. But assertions are not always given as answers to *who* or *whether* questions. It is common wisdom that belief and desire attributions are usually made to explain the agent's behaviour. It is natural to assume that an explanation is an answer to a *why* question. An explanation of why an event happens consists (typically) in stating a number of salient factors that partly were responsible for the event. The salient factors mentioned in an explanation are normally called the *cause(s)* of that event. On the assumption that belief and desire attributions are normally made to explain behaviour, what needs to be explained is a certain act or a kind of behaviour of the agent. This act can be explained in different ways, both physically, psychologically and intentionally. By means of attitude attributions we give an *intentional explanation*. When the question *Why P?* is asked it is presupposed that *P* is true. The question is asked because *P* was not expected. *P* is abnormal (with respect to a knowledge state). It is natural if something abnormal occurs that this abnormality can be explained by another abnormality. The answer to the *why* question states this other abnormality. On this reasoning, what is asserted by an attitude attribution used to explain behaviour is that the agent has abnormal attitudes. The most straightforward reason for why a certain belief is abnormal is that this belief is abnormal for *us*, the belief is different from what *we*, the participants of the conversation, believe. Normally, not everything mentioned in a sentence is really asserted. What is asserted is that part of the sentence that stands in focus. That part of the sentence that does not stand in focus is *presupposed*. For our case this means that only that part of the embedded clause of the belief attribution that stands in focus is abnormal, and the rest is presupposed. Thus, if it is assumed that the new information given in a sentence like *John believes that it stopped raining* is that John believes that it doesn't rain (anymore), it is presupposed that John believes that it was raining. Why can this be presupposed in a conversational situation where nothing has been said about the beliefs of John? The reason is that *we* believe already that it was raining, and that it is implicitly assumed that the agent shares our beliefs. This implicit assumption makes sense because agents tend to represent their environment adequately.[248]

These informal remarks cannot, of course, replace a formal way to account for the intuitions. After all, we are doing formal pragmatics in this chapter, and there is no pragmatic wastebasket available anymore. Although it is certainly possible in our two-dimensional framework to account for the intuition that (17) also presupposes that in fact it was raining, I'm not sure what is the best way to do that such that we don't get involved into binding problems.[249]

In chapter 3 we have discussed the problem of intentional identity; the problem that pronouns or short descriptions used in the embedded sentence of a belief attribution to one agent can take as antecedent an indefinite used in the embedded clause of a belief attribution of another agent. With presuppositions we have a similar problem, as can be illustrated by the following example due to Heim (1992):

(18) John believes that his parents are gone,
and Mary believes that *her* parents are gone, *too*.

---

[248] See also Heim (1992). Geurts (1995) argues that a sentence like *John believes that it stopped raining* presupposes out of context that it used to rain. He tries to account for the invited inference that John also believes that it used to rain by an explanation similar to the one I have given above.

[249] Fauconnier (1984) and Zeevat (1992) propose that, out of context, (76) presupposes that it was raining, *and* that John believes it was raining. That, I think, is unproblematic, but we have seen above that a similar proposal for (14) might lead us into binding problems. Heim (1992) and Geurts (1995) claim that the presuppositional part associated with *stop* is scope sensitive (or something like that), and thus can have scope over the belief predicate. But then it has to be explained why John also believes that it was raining.

It is well known that the presupposition associated with the particle *too* is focus dependent. If 'F' is the focus marker, the presupposition associated with *her*$_F$ *parents are gone* is not the same as the presupposition associated with *her parents*$_F$ *are gone*; only the first presupposes that someone else's parents are gone. Still, the second conjunct of (18) does not presuppose that Mary believes that John's parents are gone, although that would be predicted by our presupposition-rule if the article *too* takes scope only over the embedded clause of the belief attribution. I think that the second conjunct can even be true, if Mary has no beliefs about John's parents at all.[250] But then, Karttunen & Peters (1979) proposed, rightly I think, that the presupposition associated with particles like *also*, *too*, and *even* depends not only on what word stands in focus, but also on the *scope* of the particle. In our case this means that the particle *too* can have scope over the belief predicate. In that case we would predict that Mary believes that someone else's parents are gone. Intuitively, however, it seems that the use of *too* in the second conjunct is fine, because in the first conjunct it is stated that *John* believes that his parents are gone. How can we account for that intuition?

The crucial observation, I believe, is that the second conjunct of (18) can be appropriately uttered only if the noun phrase *Mary* has topical accent.[251] It seems that when the noun phrase that denotes the agent the belief is attributed to has topical accent, the presupposition of the embedded clause can be satisfied if it is common knowledge that some other agent believes or accepts the presupposition of the embedded clause. Thus, what is presupposed by the second conjunct of (18) is, I believe, that someone different from May believes that his or her parents are gone. This presupposition, of course, is satisfied in the context of interpretation of the second conjunct, and the use of *too* is thus appropriate.

Just as the asymmetry problem for intentional identity attributions had its *de re* variant, our presupposition problem in attitude contexts has it's *de re* variant, too. Consider the following example, again due to Heim (1992):

(19) John: I am already in bed.
Mary: My parents think *I* am *also* in bed.

Assuming that the demonstrative pronoun *I* has focal accent, that the noun phrase *My parents* has topical accent, and that *also* has scope over the belief predicate, I would predict that Mary presupposes that someone different from Mary's parents believes that someone different from Mary is already in bed. But this phenomenon has a straightforward explanation when it is assumed that presupposition is a propositional attitude, and that what one presupposes one normally also believes. In our case, when Mary asserts her sentence, she presupposes that John is already in bed. Assuming that others normally speak the truth, it can be assumed that she also believes that John is already in bed. But that is enough to satisfy the presupposition associated with Mary's utterance.

4.8 Anaphoric presuppositions

In § 4.2 we started with presuppositions associated with referential expressions, *expressive presuppositions*, but from § 4.6 on we concentrated only on presuppositions for which the reference-contexts are irrelevant. We assumed that what is presupposed and what is asserted by a sentence can be determined relatively independently of each other, that both can be represented by a proposition, a set of possible worlds, and that what is asserted can be determined in a context-independent way. But obviously what is asserted by a sentence in which referential expressions are used depends on the context in which the sentence is uttered. Thus, the existence of context dependent expressions seems to be problematic for

---

[250] Note that for this reason the modal subordination account discussed in chapter 3 can also not account, or at least not in a straightforward way, for the presupposition problem corresponding to the Hob-Nob case as exemplified by (18).

[251] Roughly, the accent Jackendoff (1972) calls B-accent.

the two-dimensional strategy, because we cannot determine the proposition expressed in a context-independent way. This was recognised by proponents of the above sketched account. They did not claim that we can determine the proposition expressed by a sentence in a context independent way, but only that given a reference-context we can determine the proposition expressed by a sentence. All we need to know is what function from reference-contexts to propositions is denoted by what is asserted by the sentence. The rise of dynamic semantics doesn't make this really more complicated. For the two-dimensional analysis of presuppositions to work, we don't need to know what actually is asserted by the sentence, all we have to know is what would be asserted by the sentence in contexts for which the ccp associated with the assertion of the sentence is defined. So, we just have to be able to determine the ccp associated with what is asserted by an utterance.

In the two-dimensional framework that I sketched above I assumed that for presuppositions it doesn't matter that the reference-contexts change during a conversation. And indeed, what has the change of a reference context to do with whether or not a context satisfies the presupposition of a sentence like *John is a bachelor*? Obviously, it are only the indices of a context that count for such a case. On the other hand, we have seen that reference-contexts are relevant for the analysis of incomplete definite descriptions like *the table*, or *the man*. The kind of examples for which Strawson's (1950) notion of presupposition as definedness condition made sense, and for which his criticism of Russell (1905) was most convincing, were cases where the presuppositional trigger was intuitively used in an anaphoric way. In these cases the descriptions are referentially used. The triggered presuppositions do not in the first place put constraints on what is assumed about the subject matter on conversation, but primarily on what is assumed about the conversation itself. Kripke (ms.) showed us that there are a lot more cases of presuppositions where the reference-contexts are crucial for an appropriate analysis.

Kripke concentrated his discussion only on a subset of expressions that are traditionally treated as presupposition triggers. He considered triggers like additive focus particles like *too*, *also*, *either* and *again*, clefts and pseudo-clefts. It is commonly assumed that additive focus particles have no effect on what is asserted, only on what is presupposed. On the assumption that except for incomplete definite descriptions it are only the indices that are relevant for the analysis of presuppositions, it seems we are obliged to say that these triggers give rise to *existential* presuppositions. And indeed, that is the way they are treated by for instance Karttunen & Peters (1979). On the traditional picture it is assumed that if a speaker makes a presupposition by using a sentence although the truth of this presupposition is not yet established in the conversation, the hearer will always be able to accommodate his context of interpretation such that the presupposition will be satisfied in this context. If the presuppositions associated with the above mentioned items were existential presuppositions, this would have the consequence that an utterance in which such a trigger occurs would be very easily appropriate in a context, and that the presupposition could also be very easily accommodated. Kripke showed, however, that these predictions are wrong. He observed that the following sentences are out of context inappropriate because of presupposition failure, although on the traditional account the presupposition of the consequent should be satisfied:

(20) *If John Smith walked on the beach last night,
then it was Betty Smith who walked on the beach last night.
(21) *If Sally opposed his tenure, it was Susan who opposed.

On the other hand, he noted that sentences like:

(22) ?Tonight *Sam*$_F$ is having dinner in New York, too.
(23) ?*Priscilla*$_F$is eating supper, again.

are difficult to interpret out of context. The same phenomenon occurs with anaphora and incomplete definite descriptions. Also (24) and (25) are difficult to interpret out of context

(24) ?He is a linguist.
(25) ?The man is sick.

So, in case of (22), (23), (24) and (25) accommodation doesn't seem to happen. Kripke concluded that the additive focus particles do not give rise to existential presuppositions that can be modelled by a set of possible worlds. Instead, he suggested that those particles induce an *anaphoric presupposition* that can only be satisfied if its antecedent is *explicitly mentioned* in the conversation. Obviously, we should conclude the same for pronouns and incomplete descriptions. With respect to clefts, Kripke was more cautious. Although the traditional account didn't seem to be appropriate, he resisted the temptation to claim that also clefts are anaphoric. The traditional analysis is inappropriate because neither an existential, nor a uniqueness presupposition (*It are John and Marry who opposed*) is asked for. What a cleft like *It is a who is P* presupposes, he suggested, is that it is an open issue who is P. Moreover, what is asserted by the cleft is an exhaustive answer to this question.

It is sometimes assumed that anaphoric presuppositions show that presuppositions cannot be accounted for in a two-dimensional way, because the presuppositions associated with (22)-(25) are not primarily about the subject matter of conversation, but about the conversational situation instead. If a speaker wants to use (24), for instance, it is required that there is a unique most salient male individual available for reference by the pronoun *he*. Thus, the presupposition associated with (24) should express this information. But why should it not be possible to express this information in terms of a proposition? Facts about the conversation are also facts about the world, and some propositions might contain only *meta-discourse* information, just as others contain primarily *meta-linguistic* information.

In chapter 2 we have already seen how to account for the 'anaphor'-like character of definite terms; there should be a unique most salient individual that satisfies the descriptive content of the description in every possibility of the context. The speaker says something about this object, but does not explicitly assert that there is such an object. Of course there are cases where we can interpret what is said by a sentence like *The S is P* although it is not yet established that there is one most salient S in every possibility of the context. First, we have cases where it is assumed that the predicate of the definite has at most one denotation in every possibility of the context, like *The king of Buranda*. Second, there are cases where the predicate might have none, one or more instantiations, like if I say *My sister studies history of art* out of context. In the latter case, the context can easily be accommodated such that it's assumed that I have only one sister. However, if it is clear that in every world in the context there are more S's, it seems impossible to appropriately state *The S is P* out of context.[252,253]

If we want to account for the above intuitions we have to put further constraints on accommodation; mere consistency and acceptability is not enough. For definite descriptions such a constraint can, I think, easily be given. If P is a potential presupposition of clause B in sentence U in context K, and is of the form *the A*, then P can only be appropriately accommodated to K if and only if $P \in f_D(B, K, U)$, K does not entail that there is no unique most salient A, and there is an S in K such that in all possibilities of S it holds that there is a most salient A.

---

[252] In both cases the definite term seems to be used attributively. Thus, if a description is attributively used, its associated presupposition can normally be accommodated, whereas if a description is referentially used it normally cannot.

[253] Sometimes, however, definite terms of the form *the S* can be used out of context although it is clear that there are more S's. Although John's car has more than one back tire, the following discourse due to Heim (1991) seems to be appropriate *John came too late for the meeting. He had a problem with his car, his back tire cracked.* I'm not sure why.

Obviously, to handle anaphoric dependencies across sentential boundaries, a context cannot only contain information about the subject matter of conversation. We saw already in chapter 2 that CCT is a theory that can account for such anaphoric dependencies because it also contains information about the discourse itself. Indeed, it has been suggested by Van der Sandt (1992) and Sæbo (1992) that CCT is the appropriate framework to account for the presuppositions triggered by additive focus particles. I agree with them, but I don't believe that the way they handle these presuppositions is entirely successful. They propose that additive focus particles give rise to anaphoric presuppositions that come with a free variable. An assertion with such a focus particle is only appropriate if this variable is bound in the context. This proposal goes a long way in explaining the facts, but I don't think it's a complete success. Kripke pointed out that a sentence of the form *aF is P, too* does not so much presuppose a specific presupposition of the form 'x is $P \wedge x \neq a$' where *x* is a variable bound in context, but that we infer from such a sentence that *a* is distinct from *everybody* of whom it is established that he or she is P, and that it is established that at least somebody is P. This more complicated presupposition is needed to account for the fact that the consequent of a sentence like

(26) If Herb and his wife both come to the party, *the bossF* will come too.

invites the inference that neither Herb nor his wife is the boss. The anaphoric account by itself does not predict this, we need something more. Something different is the case for clefts. As mentioned before, what seems to be required for an appropriate use of a cleft like *It is N who is P* is that it is presupposed to be an open issue who is the P and that this issue is a salient one in the conversation. But because we can easily interpret cleft sentences out of context, it seems that we are able to accommodate the context such that this appropriateness condition is met.

Given Groenendijk & Stokhof's (1984) work on question-answer pairs, it is easy, I think, to state what is asserted and what is presupposed by a cleft sentence. Groenendijk & Stokhof (1984) propose that a question denotes its set of complete true answers. If John, Mary, and Sue came to the party, the complete answer to the question *Who came to the party?* is that John, Mary, and Sue came to the party, and no-one else. Groenendijk & Stokhof argue that if someone answers this question by saying *John, Mary, and Sue came to the party*, he or she really means the same in these circumstances as *John, Mary, and Sue came to the party, and no-one else*. A more explicit way to answer the question with the implication that only John, Mary, and Sue came, is to use the following cleft construction: *It were John, Mary and Sue who came to the party*. This suggests that what is asserted by a cleft like *It is N who is P* is that N is P, and that *only* N is P. Building on the work of Groenendijk & Stokhof, Von Stechow (1990) defines the exhaustiviness operator *only* as follows:

$$\text{only}(Q)(P) \text{ is true iff } \neg\exists P'[Q(P') \,\&\, P \neq P' \,\&\, \forall x[P'(x) \rightarrow P(x)]] \quad \text{if Q is a positive quantifier}$$

$$\neg\exists P'[Q(P') \,\&\, P \neq P' \,\&\, \forall x[P(x) \rightarrow P'(x)]] \quad \text{if Q is a negative quantifier}$$

The effect of this definition is that if Q is a positive quantifier, P is the smallest set satisfying the generalised quantifier Q, and if Q is a negative quantifier, P is the largest set satisfying the quantifier. For a definition of positive and negative quantifiers, see Stechow & Zimmermann (1984). Important is that if clefts are interpreted in terms of *only* it can be explained already why (20) and (21) are inappropriate. In *If Sally opposed his tenure, it was Susan who opposed*, the consequent cannot be true in case the antecedent is presupposed, and Sally is not Susan.

With respect to the presupposition, we can assume that the cleft *It is N who is P* presupposes that the question who is P is the most salient topic of conversation. It is reasonable to assume that every assertion is an answer to an explicit or implicit question,

and that the question the assertion answers is the topic of the assertion. If a context is represented by a set of possible worlds, a question denotes (pragmatically) the set of possible true answers as far as is consistent with the context. In Groenendijk & Stokhof (1984) this is modelled by a partition of the context, where every element of the partition is a possible complete answer to the question. It can be the case, of course, that in a conversation more topics are under discussion. In that case, the context will be partitioned in more than one way. However, one of those topics, modelled by a non-trivial partition of the context, will be more salient than the others. If we assume an extensional variant of modal logic and follow Groenendijk & Stokhof again, we can say that the implicit question induced by the cleft construction itself, $\lambda N \lambda P \lambda w, w' \forall x[P(x)(w) \leftrightarrow P(x)(w')]$, must induce (maybe after accommodation) the most salient non-trivial partition of the context. To account for the topic of a sentence, we might say, for instance, that every formula that denotes what is asserted comes with a free variable. This free variable is interpreted in every possibility of the context by the set of partitions of the context, each representing a topic of conversation, ordered by salience.

Now for the additive focus particles. I want to propose that the potential presupposition triggered by a sentence with a proper name *a*, and an additive particle like *a F is P, too* presupposes that there is a salient individual who is P, and that for all salient individuals for which it is established that they are P, they are not identical with *a*. In this way I propose to account for the difficulty to interpret (22) and (23) out of context. Although the presupposition can be accommodated, it is not at all clear out of context to which objects or events we must attribute the relevant property. Sometimes, however, it is clear to what object we must attribute this property. As we saw above, Kripke (ms.) argued that the invited inference of the consequent of (26) is that neither Herb nor his wife is the boss. This is predicted because, given the antecedent both Herb and his wife are among the explicitly mentioned objects in the context of interpretation of the consequent who come to the party. Here is a somewhat different example

(27) *John* will leave the party at ten o'clock, and *Jill* will leave early, too.[254]

The second conjunct of this sentence invites at least the inference that John will leave early. I predict that the second conjunct presupposes that some of the very salient individuals that are not Jill will leave early. Uttered out of context, only John is a mentioned object. It follows that the only way to accommodate the context of interpretation such that the second conjunct is appropriate in its context of interpretation is to assume that leaving at ten o'clock is early. Because this is not inconsistent with the information given in the first conjunct this is what actually will happen.

Still, this kind of inference is not very easy, as shown by the following pair of discourses of Kamp & Roßdeutscher (1992):

(28) Everyone in the Gertraudenkrankenhaus remembers some *child's* cure by an intern of a blood disease. Now an intern has once again cured a *patient* of some pernicious disease.

(29) Everyone in the Gertraudenkrankenhaus remembers some *patient's* cure by an intern of a blood disease. Now an intern has once again cured a *child* of some pernicious disease.

Assuming that *again* gives rise to a similar presupposition as *too*,[255] it is predicted that (28) is correct. What about (29)? Kamp & Roßdeutscher observe that (29) is difficult to interpret. But assuming that the reasoning for (28) was correct, accommodation is possible. By accommodation it is predicted that for everyone there is at least a child cured by an intern of a blood disease, so that the patient cured by an intern of a blood disease that the

[254] After an example of Kripke (ms.), attributed to Partee.

[255] In at least the reading of *again* that is relevant here.

first conjunct talked about was a child. The difference between (28) and (29) can perhaps be explained by the fact that for (28) the presupposition of the second conjunct is already satisfied in its context of interpretation, while for (29) we have to make an additional inference.

4.9 Presuppositions in quantified contexts: the binding problem

In sloppy terms we might say that in Karttunen & Peters (1979) it was assumed that a clause of the form ∃xA gives rise to the following set of potential presuppositions:

$f_p(\exists xA, K, S) \quad = \quad \{\exists xC \mid C \in f_p(A, K, \partial S)\}$

So, it is predicted that the sentence

(30) Someone managed to succeed George V on the throne of England.

presupposes that

(31) Someone (would have) had difficulty in succeeding George V.

But the problem is that (30) is odd, although the predicted presupposition (31) is trivially satisfied, and [(30)](K) is non-empty.[256] What is needed is to make sure that the one who satisfies the presupposition must be the same as the one who verifies the assertion, which is not guaranteed by the two-dimensional formalism given by Karttunen & Peters (1979). This problem is known as *the binding problem* for two-dimensional accounts of presuppositions. As Karttunen & Peters also note, the problem is not easy to solve because the sentence can also figure as the antecedent of a conditional.

It is sometimes assumed that the binding problem is not limited to presuppositions triggered under the scope of indefinites. It is claimed that the problem comes up with presuppositions that are dependent on any quantifier. But that is a mistake. Although you might disagree with Karttunen & Peters' predictions for those cases, for the quantifiers that they consider besides indefinites, the binding problem does not arise.[257]

Karttunen & Peters suggested that it is the assumption that we should represent what is asserted and what is presupposed by separate propositions or logical forms that is responsible for the binding problem. Indeed, it is this assumption that is given up by most authors (Heim (1982), Van der Sandt (1992), Beaver (1993), Krahmer (1995)) dealing with the binding problem. It is now commonly assumed that presuppositions cannot be handled as something on top of semantics. What is asserted and what is presupposed should be more closely related with each other than proposed by the two-dimensionalists.

The above authors suggest different solutions to the binding problem, but all of them suggest to use one of the traditional alternative frameworks to the two-dimensional one to account for presuppositional inferences. Heim (1983) and Beaver (1993) proposed to use the satisfaction approach without cancellation by informativity, Krahmer (1995) suggested

---

256 Except, of course, if we represent the assertion made by (30) by ∃x[diff_to_succeed(x, gv) ∧ succeed(x,gv)]. So let us suppose we don't.

257 It would arise if it was claimed that for instance *most men managed to open the door* would presuppose that *most men would have had difficulty in opening the door*. But Karttunen & Peters could claim that such a sentence would give rise to a universal presupposition. In that case, at least no binding problem would arise.

to use the partial semantics of Peters (1977), and Van der Sandt basically proposes (an unusual variant of) the Russellian analysis.[258]

Heim (1983) suggested that the binding problem can be solved in a theory like CCT, if satisfaction is not related to worlds, but to assignments, too, if a context is represented by a set of world-assignments pairs. Thus, if *A* is an atomic clause and *P* the presupposition of *A*, the local context of interpretation S satisfies the presupposition of *A* iff S © [[P]] (S) iff $\forall$<g,w> $\in$ S: $\exists$h $\supseteq$ g:<h,w> $\in$ [[P]] (S). The most obvious problem with this proposal was that a sentence like *A man loves his cat* represented by '$\exists$xMan(x) $\wedge$ ∂(ιyCat_of(y,x)) $\wedge$ Love(x,y)' is predicted to presuppose that *every* man has a cat.

I have no serious difficulties with the (different) ways Van der Sandt, Beaver, and Krahmer try to account for the binding problem posed by a sentence like *A man loves his cat* in such a way that the unwanted prediction of Heim is avoided. On the other hand, I *do* have problems with their predicted readings or presuppositions for ordinary quantified sentences. Moreover, for reasons discussed in § 4.4, it seems to me that both the pure satisfaction account and the 'partial account' are much too rigid.

In the two-dimensional account presupposition projection can be defined case-wise, as determined by constructions or even lexical items, without thereby making any predictions about the other cases.[259] It can be expected that in this framework we can come up with reasonable presuppositions for quantified sentences. So, if we can solve the binding problem for two dimensional frameworks, it might be possible to handle quantified sentences appropriately, too. This is the main reason why I believe it's worth trying to solve the binding problem in a two dimensional framework. The other reason is that I believe it's interesting in its own right to see whether the binding problem can be solved in this framework.

Before I will present my own proposal, let me first explain why I'm not satisfied with the predicted readings or presuppositions of quantified sentences of Beaver, Krahmer and Van der Sandt. Beaver (1993) predicts that all quantified sentences give rise to existential presuppositions. I don't have any counterexample to this, but I believe that the presupposition is too weak. It is natural to assume that the oddity of a sentence in which a presupposition trigger occurs can be explained by presupposition failure. However, although an out of the blue utterance of

(32) Every German loves his Buick

is odd, Beaver cannot explain this oddity by means of presupposition failure.[260] Where Beaver's predictions are too weak, Krahmer's (1995) predictions seems to be too strong. He predicts that the above sentence can only be true if every German has a Buick. But intuitively, that's not required for the sentence to be true. It seems that the statement only

---

[258] What Heim (1983) proposed is not completely clear, the approaches of both Beaver and of Van der Sandt have some of their seeds in Heim (1983). From other work of Heim it is clear, however, that she seems to prefer the pure satisfaction account to the account favoured by Van der Sandt.

[259] On the other hand this limits the predictive power of the approach; for each new construction a new presupposition projection rule has to be defined.

[260] In Beaver (1995) it is suggested to explain the oddity of such sentences by other means, that have nothing to do with presupposition triggers occurring in such sentences. What happens in such cases, according to Beaver, is that for such sentences we accommodate that a certain set of individuals is topical, and that the sentence is about that set. I am sympathetic to this proposal, and indeed something like this I will propose too, but I'm not convinced by the claim that this accommodation has nothing to do with presuppositions. I'm happy to agree that topics can be accommodated and that topicality can explain why certain sentences in which a presupposition trigger occur are felicitous and others not, but this felicity still has to do, I think, with whether or not the calculated presupposition is satisfied.

makes a claim about Germans who have a Buick. Indeed, something like this is predicted by Van der Sandt (1992). He claims that the sentence is *semantically ambiguous* between

(33) Every German has a Buick and loves it. and
(34) Every German who has a Buick loves it.

Thus, if (32) has reading (34), Van der Sandt would say that it is asserted that (32) is only about Germans who have a car. On first sight that seems to be pretty good. But now consider an utterance of (34) in a context in which the assertion of

(35) There are fifty Germans of whom most do not have a Buick

is accepted just before. By the acceptability rules (an assertion should be informative and consistent with respect to its context of interpretation), the first reading is ruled out. So Van der Sandt would predict that a hearer infers that (32) has reading (34). However, as noted by Beaver (1993), this doesn't seem to be the case. In the context of (35), uttering (32) would be odd although claiming (34) in this context is perfectly all right. It seems that with (32) we cannot *assert* that Every German who has a Buick loves it.

I claimed that Van der Sandt's approach towards presuppositions is an unusual variant of the Russellian strategy to explain presuppositions away by scopal means. That's not the way things are presented in Van der Sandt (1992). There it's claimed that presuppositions are accommodated, if not bound. What is special about Van der Sandt's treatment of accommodation is that, just like in Heim (1983), accommodation is not limited to the main level. Besides global accommodation as it is motivated in § 4.3, it is claimed that there is also something like *non-global accommodation*. If the presupposition is not globally accommodated it is said that the utterance has no presuppositional reading.[261] It is also claimed that if the presupposition trigger occurs in a doubly embedded position, the presupposition can be accommodated not just in two, but in three different contexts of interpretation. Accommodation in an intermediate context is known as *intermediate accommodation*. Obviously, this has the same effect as saying that all presuppositional expressions can stand in non-trivial scope relations with logical operators. But if we forget about binding, Van der Sandt's account is still not just a notational variant of the traditional Russellian approach towards so-called presupposition triggers. This is because if a presupposition is triggered in syntactic position A, it can be accommodated to every syntactic position B accessible for A in standard CCT. In particular, presuppositions triggered in the consequent of a conditional or in the scope of a quantified sentence can be accommodated to respectively the antecedent of the conditional or the restrictor of the quantifier. That's why Van der Sandt comes to the prediction that in the context of (35), (32) really means (34).

What should we think of non-global accommodation? For global accommodation there exists an intuitive motivation. The hearer learned that the context of interpretation that the speaker had in mind for the whole utterance contained more information than the hearer assumed before. What could be the intuitive motivation for non-global accommodation? I find it hard to think of one. Local accommodation is normally not needed, because what is presupposed can also be entailed, and intermediate accommodation gives rise to empirical problems. With Rooth (1992), Beaver (1992) and Von Fintel (1995) I believe that the domain of quantification is not automatically restricted by the presupposition of the scope of the quantified sentence by a process like non-global accommodation. Instead, like Rooth and Von Fintel, I will assume that the representation of the assertion contains a free variable for the domain of quantification, and the calculated presupposition of the quantified sentence as a whole puts a constraint on this domain.

---

261 On this story it's not clear why the utterance should have a *presuppositional* reading in case the presupposition is accommodated to the main context.

I argued above that with a sentence like *Every German loves his Buick* we are only talking about German's who have a Buick. We have seen that by assuming the existence of intermediate accommodation it is possible to predict this, but also that it's at least not trivial to motivate non-global accommodation and that it also over-generates in that it sometimes gives rise to the wrong predictions. Note that according to the two-dimensional account of presuppositions there is no such a structural operation as non-global accommodation. And indeed, I consider this to be a virtue of the account.[262] But then the question arises how we could come to the prediction that *Every German loves his Buick* is a statement made only about Germans who have a Buick? It is here that the claim made in chapter 2 that quantifiers are anaphoric becomes relevant. There I claimed that every determiner denotes not a two, but a *three* place relation between sets. One of those sets is a salient context set that is anaphorically picked up. I claim that *Every German loves his Buick* gives rise to the presupposition that every German has a Buick, but that this presupposition only functions (i) to pick up the most salient context set that satisfies this presupposition, and (ii) to introduce a function from individuals to the set of Buicks they own. It is then with respect to this picked up context set that the assertion is made. So, just like Beaver claims that ordinary presuppositions help to select the right context of interpretation, I claim that the presuppositions of quantified sentences help to select the right context set that figures as the domain of quantification.[263] The view that quantified sentences give rise to universal presuppositions I share with Von Fintel (1995). I will assume that presuppositions that help to select context sets cannot be cancelled. The reason is that these presuppositions are, just like pronouns and (other) incomplete descriptions, expressive presuppositions; the reference-contexts are relevant for determining whether the presupposition is satisfied.

I will assume that the presupposition of a sentence whose assertion is represented by $\forall y_z(C,D)$ will be of the form $\Pi^x{}_y(A,B)$, and that this latter formula will be interpreted in such a way that variable y will afterwards refer to the most salient context set in which all A's are B's. This will be guaranteed by the following interpretation rule:

$$
\begin{aligned}
[[\Pi^y{}_x(A,B)]](S) = \quad & \{<g'[^x/_{\hat{x}}|A \wedge B|^g_h{}_y],h',w> |\ <g,h,w> \in S \ \&\ \exists \delta \in \|y\|^{g,h,w} \ \& \\
& \{d \in \delta \mid [[\hat{x}A]](\{<g,h,w>\})(d) \neq \varnothing\} \subseteq \\
& \{d \in \delta \mid \exists h':<g',h',w> \in [[\hat{x}(A \wedge B)]](\{<g,h,w>\})(d)\} \ \& \\
& \forall \delta' \in \|y\|^{g,h,w}: [\{d \in \delta' \mid [[\hat{x}A]](\{<g,h,w>\})(d) \neq \varnothing\} \subseteq \\
& \{d \in \delta' \mid [[\hat{x}(A \wedge B)]](\{<g,h,w>\})(d) \neq \varnothing\} \rightarrow \\
& (\delta' \neq \delta \rightarrow h(sw)(w)(\delta') > h(sw)(w)(\delta)] \ \&\ h[x] S h' \ \&\ h'(x) = \delta\}, \\
& \text{if } \forall \beta \in S: y \in dom(\beta) \ \&\ x \notin dom(\beta), \text{ undefined otherwise}
\end{aligned}
$$

For instance, I will assume that the calculated presupposition of *Every man loves his cat* is represented by the formula $\Pi^y{}_x(Man(x), \hat{y}(\iota v[Cat\_of(v,y)])(x))$,[264] and the calculated assertion by $\forall^x{}_z(\$Man(z), loves(v(z))$.[265] The presupposition introduces a function from individuals to their cats, and assigns to x the most salient set of men who all have cats.

Why does this suggestion not give rise to Beaver's problem with respect to intermediate accommodation? Here is why. If the speaker utters (35) he thereby introduces three relevant context sets/properties that can be picked up later by anaphoric means. First, the

---

262 Of course, non-global accommodation is also not part of the pure satisfaction account.

263 And thus indirectly help to select the right context of interpretation.

264 Remember that 'ιxA' is the formula short for '$\exists x \forall y[A[^x/_y] \leftrightarrow y = x] \wedge A$', where $A[^x/_y]$ is A with all occurrences of free x replaced by fresh y.

265 See §4.10 of thic chapter for a sketch how this could be calculated.

noun introduces the set of all German's, second, there is the set of fifty German's who are there, and third, there is a set of German's who are there who do not have a Buick. If those two sentences were the only sentences uttered, the anaphoric quantifier *Every German* in (32) *should* pick up one of those sets or the universe of discourse. But obviously, none of those sets satisfies the presupposition, thus none of those sets *could* be picked up. That's the reason why the utterance of (32) in the context of (34) is odd.

I believe with Von Fintel (1995) that a sentence like *Every German likes his cat* gives rise to a universal presupposition that helps to select the appropriate context set that functions as the domain of quantification. What is maybe not so clear about the proposal made by Von Fintel is that this proposal can only be worked out in a two-dimensional framework. But we know that the two-dimensional framework gives rise to *the binding problem*. Thus, I think that Von Fintel's proposal can only be helpful, if the binding problem for the two-dimensional framework can be solved.

I do agree with Karttunen & Peters that a sentence like

(30) Someone managed to succeed George V on the throne of England

gives rise to the presupposition that at least somebody (would have) had difficulty in succeeding George V. What we have to guarantee, however, is that at least one person who satisfies the presupposition must be the same as the person by which the assertion is verified. How can this be done? I propose that this can be done by assuming that indefinites presuppose that the speaker talked about a particular individual or group of individuals, and that what is asserted is about this individual or group of individuals. The presupposition of a sentence like *A man is sick* will be that the set of men is non-empty. The assertion says of one of these men that he is sick. The assertion will anaphorically refer back to the set of men 'introduced' by the presupposition. If the predicate of the sentence gives rise to a non-trivial presupposition by itself, the presupposition of the whole sentence will also be more contentful. For instance, in Karttunen's & Peters' example, *A man managed to succeed George V on the throne of England* presupposes that the speaker talked about the set of men who (would have) had difficulty in succeeding George V. Formally, I make a distinction between the existential quantifier in the presupposition, and the existential quantifier in the assertion. More in particular, I would represent the presupposition and assertion of (30) respectively by (30a) and (30b):

(30a) $\Sigma^{x}{}_{y}$(person(y), diff_to_succeed_GV(y))

(30b) $\exists^{y}{}_{z,z'}$('$person(z), succeed_GV(z))

The anaphoric variables in respectively (30a) and (30b) must already be defined. For the anaphoric variable of the assertion this is no problem, because the variable is introduced by the presupposition. With respect to the anaphoric variable of the presupposition, I assume that normally the variable is α, and thus always defined. When the context is accommodated with (30a), variable *y* is interpreted as the set of persons who would have had difficulty in succeeding George V. It follows that the presupposition 'introduced' the set of individuals the assertion can be about. Thus, also the presupposition associated with (30) is an expressive presupposition, and cannot be cancelled by informativity.

Formulae of the form $\Sigma^{y}{}_{x}(A,B)$ will be interpreted as follows:

$$[[\Sigma^{y}{}_{x}(A, B)]](S) = \{<g'[^{x}/_{\hat{x}}|\, A \wedge B|^{y}_{h}, y], h, w> \mid <g,h,w> \in S \;\&\; g \subseteq g' \;\&\; \exists d: \exists h' \supseteq h: <g',h',w> \in [[A \wedge B]](S[x:\approx d])\},$$

$$\text{if } \forall \beta \in S: y \in \mathrm{dom}(\beta) \;\&\; x \notin \mathrm{dom}(\beta), \text{ undefined otherwise}$$

I will assume that the presupposition of a sentence like *A man loves his cat* will be represented by Σ^{y}_{x}(Man(x), $\hat{y}$(ιv[Cat_of(v,y)])(x)), and the assertion by $\exists^{y}_{z,z'}$('$Man(z), love(z,v(z))). Thus, the presupposition introduces the set of men who have one cat, and a function from individuals to their cats, and the assertion says of one man of the set of men who own one cat, that he loves the cat he owns.

What about cases where a sentence like (30) figures as the antecedent of a conditional, or more generally, when the speaker is not responsible for the referent of the indefinite? That is really no problem. The presupposition associated with the antecedent of *If a man loves his cat, he will not beat it*, will be that the speaker is talking about a non-empty set of men who have a cat, and the same holds for a negated sentence like *No man loves his cat*.[266] What holds for a 'quantifier' like *no*, also holds for its adverbial counterpart *never*. For instance, a sentence like *Robin Hood never misses* presupposes that the speaker is talking about the set of events where Robin Hood shoots.

But what about cases like *Every$_x$ man loves a$_y$ woman*. The indefinite *a woman* does not refer to a particular set of individuals, rather it refers with respect to every man to a particular set of individuals. What will be presupposed is easy to see: for every man there is at least one woman. But we have seen that in the normal case indefinites are anaphorically related to their presuppositions. It seems clear that that cannot work in this case. I will propose that in this case the indefinite of the assertion will not be anaphorically dependent on the presupposition. In this case the indefinite will anaphorically refer to the set of individuals in the interpretation of α, the distinguished variable that refers to the set of all objects in the world of the relevant possibility.

To formalise this, we have to redefine the interpretation rule for $\exists^{y}_{x,z}$(A, B) given in chapter 2 into

$$[[\exists^{y}_{x,z}(A, B)]](S) = \{\langle g'[^{z}/_{\hat{x}}|\, A \wedge B|^{g}_{h}], h', w\rangle \mid \langle g,h,w\rangle \in S \;\&\; g' \supseteq g,\ h' \supseteq h \;\&\; \exists d \in \alpha^{g,h,w}_{||y||} \;\&\; \langle g', h', w\rangle \in [[A \wedge B]](S[x:= d])\},$$
$$\text{if } \forall \beta \in S\colon x,z \notin \mathrm{dom}(\beta), \text{ undefined otherwise}$$

In this interpretation rule, $\alpha^{g,h,w}_{||y||}$ is defined as follows:

$$\alpha^{g,h,w}_{||y||} = ||y||^{g,h,w}, \text{ if defined}$$
$$= \alpha(w) \text{ otherwise}$$

A similar story holds for definite descriptions. A description like *the man* will have a presupposition of the form ι^{x}y[Man(y)], and the assertive content will be $\lambda Q{:}\iota^{y}z$[Man(z)] $\wedge$ Qz. The presupposition assigns to variable *y* in each possibility a singleton set, where the element of this singleton set is the most salient man in that possibility. This most salient man is taken up again in the assertion. But just like for indefinites, this is not the case when the definite description stands in the scope of a quantifier. In that case it behaves in the same way as the indefinite.

4.10 Formalisation

After this rather informal discussion, let me now formalise an extensional part of what I have discussed above.

---

[266] I don't treat *no* as a quantifier, but think of it as the combination of *not* and *a*.

The grammar

The grammar is organised as follows: First I assume that there will be a syntax formalism that builds up the possible syntactic trees of the sentence by which the utterance is made. Let A' be an unambiguous syntactic tree of sentence A, S a formula, and K a context. Then I define $f_m(A')$, what is asserted by A, $\{A'\}(K)$, the semantic interpretation of A in K, $\mathbf{A}(S)$, the set of acceptable contexts for S, $f_p(A',K,S)$, the potential presupposition of A' with respect to K and S, and finally $[A'](K)$, the pragmatic interpretation rule for A'. The connectives are treated syncategorematically.

If A is a formula, $\mathbf{A}(A)$ gives the set of acceptable contexts with respect to informativity for $[[A]]$ recursively defined as follows:

$[[\Delta A]] = \{<g,h,w> | \{<g,h,w>\} \models_S A\}$

$\mathbf{A}(Px_1...x_n) = \{S | \forall x_i: 1 \leq i \leq n: \forall \beta \in S: \|x_i\|^\beta$ is defined & $S \not\subseteq [[\Delta Px_1...x_n]]$ & $S \subseteq / [[\Delta \sim Px_1...x_n]]\}$

$\mathbf{A}(x_1 = x_2) = \{S | \forall x_i: 1 \leq i \leq 2: \forall \beta \in S: \|x_i\|^\beta$ is defined & $S \not\subseteq [[\Delta x_1 = x_2]]$ & $S \not\subseteq [[\Delta \sim x_1 = x_2]]\}$

$\mathbf{A}(\sim A) = \mathbf{A}(A)$

$\mathbf{A}(A \wedge B) = \{S \in \mathbf{A}(A) | [[A]](S) \in \mathbf{A}(B)\}$

$\mathbf{A}(A \rightarrow B) = \mathbf{A}(\sim(A \wedge \sim B))$, if $\mathbf{A}(\sim(A \wedge \sim B)) \neq \emptyset$

$= \mathbf{A}(B)$ otherwise

$\mathbf{A}(A \vee B) = \{S | [[\sim B \wedge A]](S) \neq \emptyset$ & $[[\sim A \wedge B]](S) \neq \emptyset\}$[267]

$\mathbf{A}(\iota^x y(A)) = \{S | \forall \beta \in S: \|x\|^\beta$ is defined & $\|y\|^\beta$ is undefined & $[[\iota^x y(A)]](S) \neq \emptyset\}$

$\mathbf{A}(\exists^x y,z(A, B)) = \{S | \forall \beta \in S: \|x\|^\beta$ is defined & $\|y\|^\beta$ and $\|z\|^\beta$ are undefined & $[[\exists^x y,z(A, B)]](S) \neq \emptyset$ & $\sim \exists S': S \subset S'$ & $[[\exists^x y,z(A, B)]](S) = S'\}$

$\mathbf{A}(D^x y(A, B)) = \{S | \forall \beta \in S: \|x\|^\beta$ is defined & $\|y\|^\beta$ is undefined & $[[D^x y(A, B)]](S) \neq \emptyset$ & $\sim \exists S': S \subset S'$ & $[[D^x y(A, B)]](S) = S'\}$, for D any non-intersective determinor.

$\mathbf{A}(\partial A) = \wp(G \times G \times W)$

For the following definition of $f_p(\alpha,K,S)$ we need first to define when a potential presupposition is an element of q(K,S). I say that a potential presupposition τ is an element of q(K,S) iff τ contains no lambda terms, or free variables not bound in the context K, and if τ is neither entailed by K, but also acceptable in K, and if a τ is a definite description, then it must have in every element of $\cup K$ a unique most salient denotation:

$\tau \in q(K,S)$ iff $\lambda(\tau) = \emptyset$[268] & $K \not\models \tau$ & $fv(\tau,K) = \emptyset$ & $\exists T \in dom([[\tau]]): T \in \mathbf{A}(S)$ &
$(\tau = \iota^y x[Ax] \Rightarrow \forall <g,h,w> \in \cup K: kard(\{d \in \hat{x}|A|^g_h, y(w) | \forall d' \in \hat{x}|A|^g_h, y(w):$
$[d' \neq d \rightarrow h'(sw)(w)(d') > h'(sw)(w)(d)\}) \leq 1])$[269]

---

[267] If A and B are logically independent.

[268] Where $\lambda(A)$ gives the set of lambda expressions in A.

[269] thus *the man* cannot be simply accommodated.

The above definition made use of fv(A, K) defined as follows:

$fv(A, K) = \{x \in fv(A) | \exists g \in G(K): x \in dom(g)\}$

where $g \in G(K)$ if and only if there an h such that <g,h,w> or <h,g,w> is an element of $\cup K$, and where fv(A) gives the free variables of A:

$fv(Px_1,...,x_n) = \{x_1,...,x_n\}$
$fv(x_1 = x_n) = \{x_1, x_2\}$
$fv(\sim A) = fv(A)$
$fv(A \wedge B) = fv(A) \cup \{x \in fv(B) | x \notin av(A)\}$
$fv(D^x_y(A, B)) = \{x\} \cup \{z \in fv(A \wedge B) | z \neq y\}$, $D = \exists, \forall, \Sigma$ or Π.
$fv(\iota^x_y(A)) = \{x\} \cup \{z \in fv(A) | z \neq y\}$
$fv(\partial A) = fv(A)$

where av(A) is defined as follows:

$av(Px_1,...,x_n) = \varnothing$
$av(\sim A) = av(\partial A) = av(A)$
$av(A \wedge B) = av(A) \cup av(B)$
$av(D^x_y(A, B)) = \{y\} \cup av(A \wedge B)$, $D = \exists, \forall, \Sigma$ or Π.
$av(\iota^x_y(A)) = \{y\} \cup av(A)$

For the following rules I will assume that words like *and*, *or* and *if, then* only occur between two clauses of sentential level and that they occur already in the syntactic trees as their corresponding connectives.

If A' is a syntactic tree of a sentential clause, K a context and S a formula, then there will be a set of potential presuppositions corresponding with A' that are not yet cancelled by K and S. From this set we can determine *the* presupposition associated with A' relative to K and S.

$[\alpha]_S = [\sim[\beta]_S]_S$
$f_m(\alpha) = \sim f_m(\beta)$
$f_p(\alpha,K,S) = \{\{\tau \in f_p(\beta,K,S)^1 | \tau \in q(K,S)\}, f_p(\beta,K,S)^2\}$[270]

$[\alpha]_c = [\sim[\beta]_c]_c$ where $c \neq s$
$f_m(\alpha) = \sim f_m(\beta)$
$f_p(\alpha,K,S) = \{f_p(\beta,K,S)^1, f_p(\beta,K,S)^2\}$

$[\alpha]_S = [[\beta]_S \ @ \ [\delta]_S]_S$, for any binary connective @
$f_m(\alpha) = f_m(\beta) \ @ \ f_m(\delta)$

[270] Where $\{A, B\}^1 = A$, and $\{A, B\}^2 = B$.

$f_p(\beta\ @\ \delta,K,S) = \{\tau \in f_p(\beta,K,S) \cup\!| \{f_m(\beta) \rightarrow \pi |$
$\pi \in f_p(\delta,\{f_m(\beta)\}(K),S)\} | \tau \in q(K,S)\}$[271]

$[\alpha]_S = [[\beta]_{np},[\delta]_{vp}]_S$
$f_m(\alpha) = f_m(\beta)(f_m(\delta))$
$f_p(\alpha,K,S) = \{\{\gamma \in f_p(\beta,K,S)^1 \cup f_p(\delta,K,S)^1 \cup \{\pi(\mu) | \pi \in f_p(\beta,K,S)^2$
$\& \mu \in f_p(\delta,K,S)^2 \& \lambda(\pi(\mu)) = \varnothing\} | \gamma \in q(K,S)\}, \{\pi(\mu) |$
$\pi \in f_p(\beta,K,S)^2 \& \mu \in f_p(\delta,K,S)^2 \& (\lambda(\tau) \neq \varnothing \text{ or } fv(\tau,K) \neq \varnothing)\}\}$

$[\alpha]_S = [\text{even}[\beta]_{np},[\delta]_S]_S$ Even rule y[272]
$f_m(\alpha) = f_m(\beta)(\lambda y f_m(\delta))$
$f_p(\alpha,K,S) = \{\{\text{even}(f_m(\beta)(\lambda y f_m(\delta)))\} \cup \{\tau \in (f_p(\beta,K,S)^1 \cup f_p(\delta,K,S)^1$
$\cup \{\pi(\lambda y \mu) | \pi \in f_p(\beta,K,S)^2 \& \mu \in f_p(\delta,K,S)^2 \&$
$\lambda(\pi(\lambda y \mu)) = \varnothing\}) | \tau \in q(K,S)\}, \{\pi(\lambda y \mu) | \pi \in f_p(\beta,K,S)^2 \&$
$\mu \in f_p(\delta,K,S)^2 \& \lambda(\pi(\lambda y \mu)) \neq \varnothing\}\}$

where "even" means the same as in Karttunen & Peters (1979)

$\varphi[\alpha]_{np} = [[\beta]_{det},[\delta]_n]_{np}$
$f_m(\alpha) = f_m(\beta)(f_m(\delta))$
$f_p(\alpha,K,S) = \{f_p(\delta,K,S)^1 \cup \{\pi(f_m(\delta)) | \pi \in f_p(\beta,K,S)^2 \& \pi(f_m(\delta)) \in$
$q(K,S)\}, \{\pi(f_m(\delta)) | \pi \in f_p(\beta,K,S)^2 \& \pi(f_m(\delta)) \notin q(K,S)\}\}$

$[\alpha]_{vp} = [[\beta]_{tv},[\delta]_{np}]_{vp}$
$f_m(\alpha) = f_m(\beta)(f_m(\delta))$
$f_p(\alpha,K,S) = \{f_p(\beta,K,S)^1 \cup f_p(\delta,K,S)^1 \cup \{\tau | \exists \pi,\mu: \pi \in f_p(\beta,K,S)^2 \&$
$\mu \in f_p(\delta,K,S)^2 \& \tau = \pi(\mu) \& \lambda(\tau) = \varnothing \& \tau \in q(K,S)\},$
$\{\tau | \exists \pi,\mu: \pi \in f_p(\beta,K,S)^2 \& \mu \in f_p(\delta,K,S)^2 \& \tau = \pi(\mu) \& \lambda(\tau) \neq \varnothing\}\}$

$[\alpha]_{vp} = [[\beta]_{stop},[\delta]_{vp}]_S$
$f_m(\alpha) = f_m(\beta)(f_m(\delta))$
$f_p(\alpha,K,S) = \{f_p(\delta,K,S)^1, \{\mu(f_m(\delta)) | \mu \in f_p(\beta,K,S)\}\}$

$f_m(a^x{}_{y,z}) = \lambda P \lambda Q{:}\exists^y{}_{z,z'}(Pz,Qz)$
$f_p(a^x{}_{y,z},K,S) = \{\varnothing, \{\lambda P \lambda Q{:}\Sigma^x{}_y(Py,Qy)\}\}$, if $Qy \neq Ty$
$= \{\varnothing, \{\lambda P \lambda Q{:}\Sigma^x{}_y(Ty,Ty)\}\}$ otherwise

$f_m(D^x{}_{y,z}) = \lambda P \lambda Q{:}D^z{}_y(Py,Qy)$, for any non-intersective determinor D
$f_p(D^x{}_{y,z},K,S) = \{\varnothing, \{\lambda P \lambda Q{:}\Pi^x{}_z(Pz,Qz)\}\}$

$f_m(\text{his}^\alpha{}_y{}^x) = \lambda P \lambda Q{:}\$\iota^y{}_{y'}(Ty') \wedge Qy'$ bound pronoun

---

[271] Where $A \cup\!| B = \{A^1 \cup B^1, A^2 \cup B^2\}$

[272] The quantifying-in rule works very much in the same way.

$f_p(his^{\alpha}{}_y{}^x,K,S)$ = $\{\varnothing,\{\lambda P{:}\iota^{\alpha}{}_y[Py \wedge POS(y,x)],\lambda P\lambda Q{:}\iota^{y}{}_{y''}(Ty'') \wedge Qy''\}\}$

$f_m(his^{\alpha}{}_y)$ = $\lambda P\lambda Q{:}\$\iota^{y}{}_{y''}(Ey'') \wedge Qy''$ unbound pronoun

$f_p(his^{\alpha}{}_y,K,S)$ = $\{\varnothing, \{\lambda P{:}\iota^{\alpha}{}_y[Py \wedge \iota^{\alpha}{}_z(Male(z)) \wedge POS(y,z)],$
$\lambda P\lambda Q{:}\iota^{y}{}_{y''}(Ey'') \wedge \iota^{z}{}_{z''}(Ez'') \wedge Qy'')\}\}$

$f_m(love)$ = $\lambda Q\lambda y(Q\lambda x LOVEyx)$

$f_p(love,K,S)$ = $\{\varnothing, \{\lambda Q\lambda y(Q\lambda x Tyx)\}\}$, where T is the trivial relation

$f_m(man)$ = $\lambda x MANx$

$f_p(man,K,S)$ = $\{\varnothing, \{\lambda x Personx\}\}$,

$f_m(stop)$ = $\lambda X(\lambda y{\sim}Xy)$

$f_p(stop,K,S)$ = $\{\varnothing, \{\lambda X(\lambda y PAST(Xy))\}\}$[273]

$f_m(John^x{}_y)$ = $\lambda Q{:}\$\iota^{y}{}_{y'}(Ty') \wedge Qy'$

$f_p(John^x{}_y,K,S)$ = $\{\{\iota^{x}{}_y[y = John]\}, \{\lambda Q{:}\iota^{y}{}_{y''}(Ty'') \wedge Qy''\}\}$

$f_m(the^x{}_y)$ = $\lambda P\lambda Q{:}\$\iota^{y}{}_{y'}(Ty') \wedge Qy'$

$f_p(the^x{}_y,K,S)$ = $\{\varnothing, \{\lambda P\iota^{x}{}_y[Py], \lambda P\lambda Q{:}\iota^{y}{}_{y''}(Ty'') \wedge Qy''\}\}$

$f_m(he^x{}_y)$ = $\lambda Q\$\iota^{y}{}_{y'}(Ty') \wedge Py'$ unbound pronoun

$f_p(he^x{}_y,K,S)$ = $\{\{\iota^{x}{}_y[Male(y)\}, \{\lambda Q\iota^{y}{}_{y''}(Ty'') \wedge Py''\}\}$

$f_m(he_y)$ = λPPy bound pronouns

$f_p(he_y,K,S)$ = $\{\varnothing, \{\lambda PPy\}\}$

$f_m(is)$ = $\lambda P\lambda x P(\lambda y[x = y])$

$f_p(is,K,S)$ = $\{\varnothing, \{\lambda P\lambda x P(\lambda y[x = y])\}\}$

$\{A\}(K)$ = $\{[[f_m(A)]](S) | S \in K\}$

Let A be a syntactic tree, then we can finally define

$[A](K)$ = $\{[[f_m(A)]](S') | \exists S \in K \;\&\; [[f_p(A,K,f_m(A))]](S) = S' \;\&\; S' \in A(f_m(A))\}$

Assume that A and B are formulae, and A' a syntactic tree:

$A \models_{d/s} B$	iff	$\forall S \in dom([[A]])$: $[[A]](S) \models_{d/s} B$	semantic entailment
$A' >> B$	iff	$\forall S \in A(f_m(A'))$: $S \models_{d/s} B$	clausal implicature
$A' \to_K B$	iff	$\forall S \in f_p(A',K,f_m(A'))$: $S \models_{d/s} B$	presupposition
$A' \Rightarrow_K B$	iff	$\forall S \in [A'](K)$: $S \models_{d/s} B$	reasonable inference

[273] Because I won't introduce events, this is all I will say about such aspectual verbs.

Chapter 5

Conditionals and belief change

5.1 Introduction

In this chapter I will discuss conditionals. There are several reasons why I want to do this. First, because I think the analysis of conditionals is interesting in its own right. Second, because the causal-information theoretic account of content that I favour relies on an analysis of causal and counterfactual relations. Third, and most important for my purposes, because a discussion of conditionals allows me to introduce some concepts and some distinctions that I believe can be used for the pragmatic analysis of some attitudes and permission sentences in the next chapter. For instance, I will introduce probability functions that will be used for the analysis of desire. The main point I want to make with respect to the next chapter is that the triviality result of Lewis (1975a) suggested that there are two conceptually different ways in which one can change one's belief state. The one is learning new information, and the other is change by action. Although different, I want to look at how those two ways can be related to each other via a construction due to Harper (1976). This construction will play an important role in the next chapter. With respect to the analysis of conditionals itself, I will argue that all conditionals state propositions and should be analysed in a similar way, but what proposition is expressed by a conditional is context dependent. Just like in earlier chapters, sometimes we need to rely on diagonalisation to be able to determine what proposition is expressed by a certain conditional. To account for the context dependence of the appropriateness of the assertion of a conditional sentence, I will suggest that maybe we should use a strict conditional account. But by using Warmbrod's (1981) analysis, this account will be closely related to the traditional analysis of counterfactuals by Lewis and Stalnaker. With respect to the causal-information theoretic account of content, I will argue that in the end even the most objective-like counterfactual relations cannot be reduced to purely non-intentional matters of fact. Belief and intention are still needed for the analysis.

5.2 The Lewis/Stalnaker analysis of conditionals

Let us assume that sentences of the form *if A then B* should have a uniform analysis. Then the question arises what the right analysis can be. The conditional *If the butler didn't do it, the gardener did* can be denied without affirming the butler's guilt. That's why the material implication isn't a good representation of natural language *if A then B*. In this respect, Belnap's (1970) analysis of conditional assertion is better. It analyses the assertion of a conditional as the assertion of the consequent conditionally on the truth of its antecedent. But because the antecedent of a counterfactual is by definition false, this analysis cannot account for an important group of conditionals. It seems we have to consider other ways the world might have been to interpret these conditionals. In other words, we have to go from an extensional to a modal analysis of conditionals. In this way we can analyse he conditional *if A then B*, as follows:

If A and .... then B

But by analysing conditionals as strict conditionals, the blank can only be filled by all maximal sets of (invisible) premises that are consistent with A. Because both *If I had shirked my duty, no harm would have ensued* and *If I had shirked my duty and you had too, harm would have ensued* can be true simultaneously, the strict conditional analysis will not do.[274] Assuming a fixed accessibility relation for the necessity operator of the strict conditional, it cannot only not account for the non-monotonic behaviour of counterfactuals

[274] Stalnaker and especially Lewis argued that the strict conditional account can also not account for evidence against the validity of contraposition and transitivity. Most authors agree with Stalnaker and Lewis, but we will come back to this.

and conditional assertions, it also does not pay enough attention to the flexibility to make the interpretation of the conditional dependent on what the speaker had in mind. The actual world, the influence of the intention of the speaker and the non-monotonicity of conditionals must be reflected somehow in the right analysis of conditionals. Lewis and Stalnaker developed an analysis of conditionals[275] in which those three requirements are met by introducing (context dependent) selection functions that take as arguments both the actual world and the antecedent of the conditional. In this way, the blank can be filled with enough but not too many additional premises. We will turn now to their possible world analysis.

In possible world semantics, a proposition expressed by a sentence is equated with the set of possible worlds in which the sentence is true.[276] If we represent conditionals as A > C, Stalnaker (1968) and Lewis (1973) gave the following analysis:

A > C is true at *w* if some A ∧ C-worlds are closer to *w* than any A ∧ ~C-worlds.

To make sense of this definition it is needed to explain what 'closer A-world to *w* than' means. Therefore an ordering relation on the accessible worlds (but let us assume that all worlds are accessible) with respect to *w*, $\leq_w$, is defined which meets the following conditions:

(a)	reflexivity	$\forall w'$: $w' \leq_w w'$;
(b)	transitivity:	$\forall w',w'',w'''$: $(w' \leq_w w'' \ \& \ w'' \leq_w w''') \Rightarrow w' \leq_w w'''$;
(c)	connectedness:	$\forall w',w''$: $w' \leq_w w''$ or $w'' \leq_w w'$;
(d)	strong centering:	$\forall w'$: $w \neq w' \Rightarrow (w \leq_w w'$ and $\sim w' \leq_w w)$.

So, the relation $\leq_w$ is a weak ordering that meets the strong centering assumption. The intuitive meaning of $w' \leq_w w''$ is that *w'* is at least as close (or similar) to w as *w''* is. The relation of similarity leads to a notion of *sphere around w*. A sphere around *w* is a set of possible worlds. Two possible worlds w',w'' are in the same sphere around *w* iff $w' \leq_w w''$ and $w'' \leq_w w'$. Note that by the strong centering assumption {w} is a distinguished sphere around *w*. The spheres can be ordered by the following $\$_w$-relation:

$\forall X \subseteq W, Y \subseteq W, \$_w(X, Y)$ iff $w' \leq_w w''$ for all $w' \in X$, $w'' \in Y$

This $\$_w$-relation totally orders the spheres relative to w. A system of spheres centred on {w} is a collection **S** of subsets of W, $\{S_i : i \in \mathbf{N}\}$, that satisfies the following conditions:

(S1) **S** is totally ordered by $\subseteq$; $(S_i, S_j \in \mathbf{S} \ \& \ i \leq j) \Rightarrow S_i \subseteq S_j$;

(S2) {w} is the $\subseteq$-minimum of **S**; $\{w\}, S \in \mathbf{S} \Rightarrow \{w\} \subseteq S$;

(S3) $W \in \mathbf{S}$, and

(S4) If there is a sphere in **S** intersecting a proposition A, there is a smallest sphere in **S** intersecting A.

---

[275] Lewis claims that the analysis only works for counterfactuals. Stalnaker argued that the analysis also applies to indicative conditionals.

[276] In this chapter we will use capitals to denote both natural language sentences and the propositions they express. I hope this will never lead to confusion.

Lets call the smallest sphere is S that gets with A a non-empty intersection, S_A. The set of most similar A worlds to *w* is now $A \cap S_A$. By using only the above constraints on models, we end up with Lewis's semantics for counterfactuals. It has two special characteristics that makes it different from Stalnaker's original semantics for conditionals. First of all if the field of $\leq_w$ is infinite, it allows for closer and closer A-world without (a set of) closest. Second, different possible worlds can be closest to *w* without being identical to each other, that is, ties are allowed. Stalnaker prohibits both possibilities by the uniqueness assumption: $A \neq \emptyset \Rightarrow \exists!w' \in A$ such that $\forall w''[w'' \in A \rightarrow w' \leq_w w'']$. This uniqueness assumption can be represented by the following two constraints on models:

(e) limit assumption: $A \neq \emptyset \Rightarrow \{w' | w' \in A \ \& \ \forall w''[w'' \in A \rightarrow w' \leq_w w'']\} \neq \emptyset$

(f) trichotomy: $\forall w', w''$: $w' <_w w''$ or $w'' <_w w'$ or $w' = w''$.

Note that by trichotomy, the equivalence classes induced by $\leq_w$ turn out to be singleton sets. Accepting the limit assumption[277] (or limiting our analysis to the finite case) and trichotomy, we can reformulate the semantics of counterfactuals in the following way: Let's call $M = \langle W, f \rangle$ a model structure in which W stands intuitively for the set of possible worlds and *f* for a function such that $\forall w \in W$, $A,B \subseteq W$: $f_w(A) \subseteq W$, $f_w(B) \subseteq W$ and satisfies the following conditions:[278]

(a) $f_w(A) \subseteq A$,

(b) $f_w(A) = \{w\}$, if $w \in A$,

(c) if $f_w(A) \subseteq B$ and $f_w(B) \subseteq A$, then $f_w(A) = f_w(B)$, and

(d) $f_w(A)$ contains at most one member.

The proposition expressed by the conditional A > B is the following set of possible worlds:

$$A > B = \{w \in W : f_w(A) \subseteq B\}$$

That is, A > C is true in *w* iff C is true at every closest A-world to *w*, or A is impossible. Conditions (a)-(c) are assumed by both Stalnaker (1968) and Lewis (1973). The first condition guarantees that the truth value of a conditional is partly dependent on the truth value of the antecedent. Condition (b), the strong centering assumption, gives already some content to the notion of similarity. On the one hand it guarantees that if the antecedent is true, the conditional behaves like material implication, so modus ponens is valid. It assures that in case the antecedent is true, it is only the actual world that counts. On the other hand, it also validates the inference from $A \wedge C$ to $A > C$.[279] Condition (c) gives the most content to the similarity function. It says that the similarity function is independent of the antecedent (or conditional) to be evaluated. Lewis and Stalnaker disagree whether or not condition (d) should be given. The principle that corresponds with this condition, $(A > B) \vee (A > \neg B)$, is known as the principle of the *conditional excluded middle* (CEM). It is proposed by Stalnaker to capture the intuition that we deny a conditional *if A then B* by *If A then not B*. Should we accept this principle? Stalnaker keeps saying *yes*, Lewis keeps saying *no*. Lewis says *no*, because assuming (CEM) makes it unclear how to account for *might* counterfactuals, and anyway, ties are needed to account for Quine's sentences

---

277 I will always make the limit assumption.

278 Or we define the selection function in terms of the similarity relation as follows:
$f_w(A) = \{w' \in A | \forall w'' \in A: w'' \leq_w w' \Rightarrow w'' = w'\}$

279 If we replaced strong centering by weak centering - $w \in A \Rightarrow w \in f_w(A)$ - this inference would no longer be valid.

(1) If Bizet and Verdi were compatriots, Bizet would have been Italian.
(2) If Bizet and Verdi were compatriots, Verdi would have been French.

How else can the fact that, intuitively, neither (1) nor (2) is true, be accounted for?

Stalnaker argued instead that *might* is epistemic and that it has normally wide scope over the conditional. Epistemic *might* doesn't say something about a single possibility, but about a whole information state, instead. So, if an information state **K** is represented by a set of possible worlds, K and a selection function, *f*, a *might* counterfactual, $\Diamond(A > C)$ is acceptable in **K** iff there is a world in K in which $A > C$ is true with respect to the selection function. But what if there is only one world left? Stalnaker argues that even in that case he can account for *might* counterfactuals. The idea is that it might be unclear what the right way is to select nearest possible worlds. We don't associate with a possible world a single set of spheres, but rather a set of sets, for each selection function one. Conditional statements are not simply true or false in a world, but true or false in a world with respect to a selection function. Given that we have a set of selection functions, F, the notion of absolute truth and falsity is then defined in terms of supervaluation. A conditional is absolutely true (false) in a world, if it is true (false) in this world with respect to all selection functions in F. This account makes it possible that some counterfactuals are neither true nor false in a world. Stalnaker suggests that this is what is going on with Quine's sentences.[280] In this way it also becomes possible to account for *might* counterfactuals: the counterfactual $\Diamond(A > C)$ is true in *w* if there is a selection function $f \in$ F such that $A > C$ is true in *w* with respect to *f*.[281]

Recently, Von Fintel (1994) argued that *only if* clauses give a good motivation to make the uniqueness assumption. To give an account of such clauses, it would be preferred to analyse them compositionally in terms of the meaning of *only* and *if*, instead of treating *only if* as one connective. The main empirical constraints is that *B, only if* A seems to mean the converse of *If A, then B*. According to a well established tradition, as in Rooth (1992), if A is in focus *Only A* is true iff A is true and all of the relevant alternatives to A are not true. It is assumed that *only A* is interpreted with respect to an invisible contextual parameter C which contains a set of propositions, and that 'only$_C$(A)' expresses the following proposition: $\{w \in W | \forall B \in C: w \in B \Rightarrow A \subseteq B\}$. The question now is how we should interpret *if A, then B* such that *B, only if A* means that for no relevant alternative D to A, *if D, then B* is true. (also) its converse.[282] First, note that in *B, only if A*, normally it is A that has focus. Suppose *if* means what the similarity approach predicts it to mean. In that case, the proposition expressed by *B, only if A* with respect to context C is true in *w* iff $f_w(A) \subseteq B \ \& \ \forall D \in C[D \neq A \rightarrow f_w(D) \not\subseteq B]$, and for no two elements of C, the one is entailed by the other. Note that in general we cannot infer from this proposition to $B > A$, because there might be a $D \in C$ such that $f_w(D) \not\subseteq B$, but $f_w(D) \cap B \neq \emptyset$. Just consider the case where $C = \{A, D\}$, $A = \{v\}$, $f_w(A) = \{v\}$, $B = \{u, v\}$, $f_w(B) = \{u, v\}$ and $f_w(D) = \{u, x\}$. Once we make the uniqueness assumption, however, it can be explained why this will not happen.[283] Indeed, Von Fintel (1994) argues that *only if* constructions gives us a good reason to make the uniqueness assumption.

---

[280] For some additional arguments for Stalnaker's position, see Stalnaker (1984, ch. 7).

[281] Van Fraassen (1974) has proved that Lewis models which satisfy the limit assumption are equivalent to a family of Stalnaker models that satisfy the so-called *regularity* condition. The idea is that if $\{\langle W, f_S\rangle | f_S \in F\}$ is a family of Stalnaker models, then we can define a Lewis model $\langle W, f\rangle$, where for every $A \subseteq W$: $f_w(A) := \bigcup\{f_{S,w}(A) | f_S \in F\}$, if the family of Stalnaker models satisfies the regularity condition.

[282] I say that it *also* means its converse because I think that *B, only if A* really means *A iff B*.

[283] Suppose $f_w(B) \notin A$, then there must be world closer to w where B is true as the closest world to w where A is true. But it is reasonable to assume that in that case there will be a $D \in C$ such that $f_w(D) \in B$,

Stalnaker (1980a) argued that also if there is in fact no penny in my pocket, although I do not know it since I did not look, the falsity of

(3) If I had looked, I might have found a penny

can be explained by giving *might* wide scope. What should be done is that in this case we don't consider all the worlds compatible with my knowledge, but only those worlds compatible with my knowledge that I would have if I knew all the relevant facts. On this quasi-epistemic reading of *might* the conditional comes out false as wanted. Important is that even assuming this quasi-epistemic context of interpretation, this account still leaves open the possibility that the truth value of some conditionals can be indeterminate. Lewis (1973) argued, however, that the use of *might* in (3) has something to do with objective chance (indeterminism), which should not be modelled by epistemic uncertainty. But once you accept the supervaluation account of Van Fraassen (1974) and Stalnaker (1980a) for conditionals, you might argue that chance should be modelled by quantifying (put the probability measure) over (a set of relevant) selection functions.[284]

Again, Lewis argued that *would* and *might* counterfactuals are true or false with respect to a world and thus analyses them distributively. Stalnaker, instead, proposed a global, purely epistemic analysis of *might* counterfactuals. According to Lewis, $\Box\!\!\rightarrow$ and $\Diamond\!\!\rightarrow$ are *two different connectives*. There is no single interpretation of the if-clause. Stalnaker instead gives a single interpretation to the if-clause, and gives the modal adverb wide scope.

Unsurprisingly, the different analyses of *might* and *would* counterfactuals by Lewis and Stalnaker have their counterparts in their analyses of probabilities of (subjunctive) conditionals. While Stalnaker prefers a wide scope analysis of probability, Lewis proposes instead a distributive analysis. If w(A) is the truth value of A in *w*, Stalnaker (1970a) determines the probability of a conditional like A > C in the following way, $P(A > C) = \sum_W P(w) \times w(A > C)$. Assuming the existence of only one selection function, in a world, a conditional is either true or false, only in an information state does it make sense to talk of probabilities. Lewis (1973), instead, allows for conditionals to have a non-trivial probability in a world. According to him, the probability of A > C in *w* is roughly the same as $P(C/f_W(A))$.

To me it seems reasonable that both kinds of probabilities do measure something relevant. It is remarkable that Skyrms' (1980a, 1980b, 1994) analysis of objective chance of C given A in world *w*, $CH_W(C/A)$, is exactly that what is suggested by Lewis to be the probability of a conditional in *w*, $P(C/f_W(A))$. For Skyrms (and Lewis (1981a)), the probability of a conditional is defined as follows: $P(A > C) = \sum_W P(w) \times CH_W(C/A)$,[285] justice is done to both epistemic uncertainty and objective chance. But still there is the question whether objective chance should not be reduced to subjective probability.[286]

---

which was ruled out by our truth definition. In our case, if $f_W(B) = \{u\}$, it is reasonable to assume that either $f_W(D)$ would have been {u}, too, or that u is a possibility not under consideration.

[284] This suggestion is appealing especially because of the result of Van Fraassen (1974) showing, under certain conditions, that Stalnaker models are Lewis models with hidden variables, and thereby making a connection between the hidden variable interpretation of quantum mechanics and Stalnaker's analysis of conditionals.

[285] We will see later that a simpler relation between probabilities and conditionals leads to problems.

[286] Actually, Skyrms (1994) determines the selection function out of epistemic principles and claims that he reduces objective chance to subjective probability.

5.3 The Ramsey test analysis

Lewis and Stalnaker analysed conditional sentences semantically in terms of selection functions. They motivated the properties of those selection functions partly by purely empirical considerations. However, it would be nice if *independent* motivation for those properties could be given. This is tried in Stalnaker (1968, 1970a). In analytic philosophy it is traditionally assumed that objective modal concepts should be reduced to subjective ones. In the spirit of this tradition Stalnaker tried to understand the content of conditional propositions, and thus to motivate the properties of the selection functions, in terms of conditional beliefs. Stalnaker's (1968) analysis of conditionals is a generalisation of a suggestion first made by Ramsey (1931) and therefore called the *Ramsey test* analysis of conditionals. Ramsey's pragmatic philosophy inspired him to reduce the meaning of sentences to beliefs. His natural suggestion was that the analysis of conditional sentences should be reduced to *conditional beliefs*. Stalnaker gives the following instructions for deciding whether you do or do not believe a conditional:

> 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 consequent is true. (Stalnaker, 1968, p. 102)

According to the above quotation, conditionals should be handled in terms of *belief revision*. If we represent a belief state by K, the Ramsey test analysis can be stated in the following way:

A > C is true (accepted) in K iff C is true (accepted) in K revised by A.

Analysing conditionals in terms of changing beliefs, connects the interpretation of conditionals with that of inquiry.

> Inquiry is the process of changing [...] acceptance states, either by interaction with the world or by interaction between different acceptance states. Methodological policies are policies constraining such changes. To have a framework for describing methodological policy, we might assume that acceptance states have two components: a set of alternative possibilities representing the inquiry's current conception of the way the world is, and a change function representing his disposition to change what he accepts in response to new information. (Stalnaker, 1984, p. 99)

The properties of selection functions can be derived if it is assumed via the Ramsey test analysis that conditional propositions are projections of the methodological policies onto the world. Suppose that <K,f> is our belief state, where K is a set of alternative possibilities and *f* the change function. Now suppose that we learn so much about the world that K consists of only one possible world. In that case, *f* will coincide with the selection function used in the *semantic* analysis of conditionals. What properties can be derived from this methodology? That depends of course on how belief revision should be analysed. But it is clear that for instance the condition '$f_w(A) = \{w\}$ if $w \in A$' can be motivated in this way: if you believe A already, you don't have to change your beliefs if you learn A. More in general, the Ramsey test analysis gives a motivation why the notion of similarity should play a part for the analysis of conditionals. If you have to change your beliefs in response of new evidence, you only want to give up those beliefs for which the new evidence gives you reason to.

Let us assume with Stalnaker that there always exists a selection function *f* that assigns to every world *w* and proposition *A* a single possible world *w'*, possibly the same as w. Suppose now that an information state is represented by <K,f>, where K is the set of possible worlds in which all the accepted propositions are true, and f is the selection function. We can now define the change (revision) function of K, C, in the following straightforward way:

$C_K(A) = \{f_w(A) : w \in K\}$[287]

If the selection function satisfies the constraints that Stalnaker (1968) argued for, it is justified to say that $C_K(A)$ is the minimal revision of K by A.

If, by the Ramsey test, belief in conditionals should be reduced to conditional belief, the analysis goes like this:

(a) $A > C$ is accepted in K iff $C_K(A) \subseteq C$

If the agent learns more and more about the world, ideally he once reaches an information state where K consists of only one possible world. In that case (a) comes down to (b):

(b) $A > C$ is true in w iff C is true in $f_w(A)$

The informal connection between belief in conditionals and conditional belief made in Stalnaker (1968) gives, arguably, a good motivation for those constraints on selection functions that Lewis and Stalnaker agree on.[288] But it does not determine at all whether or not $f_w(A)$ should contain at most one member. Suppose we don't want to be committed to the assumption of trichotomy, then the definition of $C_K(A)$ changes into

$C_K(A) = \cup\{f_w(A) : w \in K\}$

In case K consists of only one world, (a) comes down to (c):

(c) $A > C$ is true in w iff C is true in all w' in $f_w(A)$

Above I defined the global change function, $C_K(A)$, in terms of selection functions on worlds. But that is not really what the Ramsey test analysis suggests, it should rather be the other way around. The selection function on individual possible worlds should be determined by how rational agents would change their global belief state by learning new information. Stalnaker (1968) tried to explain the proposition expressed by a conditional sentence as a projection of the epistemic notion of conditional belief onto the world. There is another tradition that tried to do something similar. According to the Bayesian account, objective modal notions like *causality* and *chance* are tried to be reduced to their subjective counterparts. Stalnaker (1970a) intended to explain the properties of the selection function on individual possible worlds by making a connection between the Bayessian account of belief change and his 1968 analysis of conditionals. Moreover, by making this connection he wanted to settle the issue between Lewis and himself about the controversial Conditional Excluded Middle principle. To that we will turn now.

5.4 The Bayesian approach

Independently of the analysis of conditionals, another model of changing information states had been developed. In the epistemic *Bayesian tradition*, information states (beliefs) of agents are modelled by probability functions and the (rational) change of information states is handled by *conditionalisation*.[289]

It is normally assumed that any propositional probability function, Pr, has to obey the following three constraints for all wffs A and B:

---

[287] See Stalnaker (1984) footnote 2, chapter 7.

[288] As always, made on the assumption that we should accept the limit assumption.

[289] The classical paper on Bayessianism is Ramsey (1931), the best introduction is Jeffrey (1965).

(a) $1 \geq Pr(A) \geq 0$
(b) $Pr(T) = 1$, for any truth-conditional tautology T
(c) $Pr(A \vee B) = Pr(A) + Pr(B)$, if A and B are logically disjoint.[290]

A wff A is Pr-*valid* iff $Pr(A) = 1$ for all Pr's.
If S is a set of wffs, then A is Pr-*entailed* by S if for all Pr, if $Pr(B) = 1$ for each member B of S, then $Pr(A) = 1$.

Two kinds of justification have been given for those subjective probability functions. A logical justification of *consistency*, and a pragmatic justification of *coherence*.

It can be shown that a wff is Pr-valid iff it can be proved by any standard axiomatisation of classical logic.[291] So, the probability functions are at least logically *consistent*.

Propositional probability functions have been given a pragmatic justification of *coherence* by means of the *Dutch book theorem*. It is assumed that probability functions model a belief state of an agent, and to prove this theorem, it is assumed that the degree of belief of a proposition attributed to an individual is determined by the minimal odds at which the individual is inclined to accept a bet on the truth of that proposition. Only if the degree of belief the subject assigns to A is greater or equal to n, he is willing to bet on A at odds n/(1 - n). The Dutch book theorem establishes that no bet can be constructed such that whatever happens the individual will lose money, if the individual's degrees of belief are represented by a probability function constrained by the above conditions.

Normally, conditional probability functions are defined in terms of singular ones in the following way:

> For all propositions B, the probability of B in the new probability function Pr^*_A which is the probability function Pr as changed by the (Pr-compatible) proposition A is equal to $Pr(A \wedge B)/Pr(A)$

In other words,

> $Pr(B/A) := Pr(A \wedge B)/Pr(A)$, when $Pr(A) \neq 0$, otherwise undefined.

So in the Bayesian tradition too, the rational change of information states is defined in terms of *conditional beliefs*. Just like Stalnaker's account of conditionals was based on the analysis of minimal belief change, the analysis of minimal belief change in the Bayessian tradition also gave rise to an account of conditionals. Adams (1965, 1976) claimed that the assertability of an indicative conditional goes with its corresponding conditional probability. To have a logic of conditionals, the notion of Pr-entailment is not the most interesting one. Adams based his analysis of conditionals on the following notion of ε-entailment:

If S is a set of wffs, then A is ε-*entailed* by S iff for every $\varepsilon > 0$ there is a $\delta > 0$ such that for all Pr, if $Pr(B) > 1 - \delta$ for each member B of S, then $Pr(A) > 1 - \varepsilon$.[292]

---

[290] It is clear that those constraints presuppose a logic. It is possible to give equivalent constraints on probability functions without presupposing such a logic. See Popper (1959), Stalnaker (1970a) and Leblanc (1983).
[291] See Leblanc (1983).
[292] If S contains *n* propositions, you might think of δ as being ε/n.

Note that the resulting logic is non-monotonic. Learning new information can easily decrease certain (conditional) probabilities. Also principles like modus tollens and transitivity are invalid according to this notion of entailment. Adams claims that these principles should not be valid, not even for the analysis of indicative conditionals.

A very popular extension of rational belief change by conditionalisation was invented by Richard Jeffrey (1965), and called *Jeffrey conditionalisation*. It represents changes by inputs of the form 'the probability of B is *a*' ($1 \geq a \geq 0$). Representing the actual belief state by Pr, the revised state is represented by Pr' and defined in the following way:

For all C, $Pr'(C) = a.Pr(C/B) + (1 - a).Pr(C/\sim B)$

Note that traditional conditionalisation is a special case of it. Important is that Jeffrey stressed that this rule of the kinematics of rational belief is all that is needed. It's never needed to give up propositions that are assigned probability one. The assumption is that only logical tautologies have probability one and only contradictions probability zero. This view is defended by means of the condition called *strict coherence*. A probability function obeys strict coherence if it is coherent and there is no set of bets consistent with it such that the better might lose, and cannot win. Jeffrey, among others, has argued that by assuming strict coherence it follows that no contingent proposition should have probability one, because no rational agent would bet his head on the truth of any contingent proposition.

In the 20'th century, probability theory has been widely used in the philosophy of science to construct logics of induction or confirmation. These logics are developed to capture the inductive (causal) relations between propositions. As is well known, Popper argued that scientific theories are *not infallible*. Sometimes some propositions accepted before have to be given up because of some new phenomena. So, it cannot be that only tautologies are accepted.

If it is assumed that a belief state of a rational agent who accepts more than just logical tautologies is represented by a probability function, and that a proposition is accepted by this agent iff its probability function assigns to it probability 1, It follows that also some contingent propositions are assigned probability 1 by this probability function. In that case it makes sense to define the information state based on probability function Pr in the following way: $\Omega(Pr) = \{A: Pr(A) = 1\}$. Such an information state obeys the characteristic properties of what might be called an *acceptance state*: For all propositions A and B:

(a) if $A, B \in \Omega(Pr)$, then $A \cap B \in \Omega(Pr)$

(b) if $A \in \Omega(Pr)$ and $A \subseteq B$, then $B \in \Omega(Pr)$

(c) $A \cap \sim A = \varnothing \notin \Omega(Pr)$

However, according to classical Bayesians, a contingent proposition should never have probability 1. The problem with this view is that the conditions for being an acceptance state are no longer guaranteed if propositions are already believed if their subjective probability are higher than for instance 0.5.[293] Thus, $A \in \Omega(Pr)$ iff $Pr(A) > 0.5$. But on this assumption, we can easily find counterexamples to condition (a). Consider the following two statements about our next throw of an unbiased dice: *It will show 1, 2, 3 or 4*, and *It will show 3, 4, 5 or 6*. Their subjective probabilities are both higher than 0.5, but their conjunction certainly is not. Thus, if a sentence is accepted iff its probability is higher than 0.5, the set of accepted sentences cannot be closed under conjunction. But also condition (c) is not guaranteed. This is made clear by the (infinite) lottery paradox. Assume a lottery with (infinitely) many tickets. For each ticket *i* the probability that it will not be the winning ticket will be more than 0.5. But it is clear that this acceptance state is inconsistent,

[293] or any other real number in [0, 1).

if it were closed under conjunction. To account for acceptance, or more in general for monotonic reasoning, Skyrms (1980a) does not use the notion of probability, but that of *resilience*. Resilience is a measure of *invariance under belief change*.[294] A proposition is accepted if its resilience is greater than 1/2. Skyrms (1980a, p.152-154) proves that in this way an acceptance state obeys the three conditions above.[295]

The definedness condition on classical and Jeffrey conditionalisation has the immediate consequence that if we want to represent beliefs in terms of probability functions, it becomes impossible to analyse *counterfactual beliefs*. The reason is that by this definition, Pr becomes a partial function, Pr(A/B) is undefined when Pr(B) = 0. Unsatisfied with this partiality of Pr and motivated by his own philosophy of science, Popper (1959) gave conditions on probability functions, P, where P is always defined. This is done by taking conditional probability as basic. These functions are normally called *Popper functions*, but via Stalnaker (1970a) also known as *extended probability functions*:

Popper argued that to account for the notion of acceptance, we should give up strict coherence. To capture the notion of acceptance, more propositions than only tautologies have to be assigned probability one. By taking conditional probabilities as basic, Popper functions contain more information than standard probability functions. It contains the extra information of how one would change one's belief state by learning something that is inconsistent with one's present belief state. So, (one step) revision is built into the probability function. This extra information also captures invariance under belief change or *epistemic entrenchment*. Suppose two propositions are both believed, then one of the two can still be stronger believed than another, if the latter would be given up earlier than the former. Why is one believed proposition given up earlier than another? The reason is that the former proposition is stronger connected to other accepted propositions than the latter. Popper-functions contain information about the inductive and conceptual relations among propositions believed. If the correlation between the changing probability-values of two propositions under different counterfactual assumptions is high, it is likely that the events described by the propositions are connected with each other. The difference between logical tautologies and contingent propositions that are both accepted is then that the former will never be given up, while the latter will.

A *Popper function* or an *extended probability function* (epf) is any function, P, taking ordered pairs of wffs into real numbers which meet the following five conditions for all wffs, A, B and C:

(a) $0 \leq P(A/B) \leq P(A/A) = 1$
(b) If $P(A/B) = 1 = P(B/A)$, then $P(C/A) = P(C/B)$
(c) If $P(C/A) \neq 1$, then $P(\sim B/A) = 1 - P(B/A)$
(d) $P(A \wedge B/C) = P(A/C) \times P(B/A \wedge C)$
(e) $P(A \wedge B/C) \leq P(B \wedge A/C)$[296]

A wff A is P-*valid* iff $P(A/B) = 1$ for all epfs P and propositions B
If S is a set of wffs, then A is P-*entailed* by S if for all epfs P and propositions B, $P(C/B) = 1$ for each member C of S, then $P(A/B) = 1$.

---

[294] Just like the notion of epistemic entrenchment that we will discuss later.

[295] Using non-standard probability assignments, where propositions with ordinary probability 0 are now given an infinitesimal probability to make conditionalisation always defined. See also Pearl (1994).

[296] We have seen that proponents of strict coherence used non-standard probability assignments to assure that conditionalisation is always defined, while Popper simply took conditional probabilities to be primitive to assure the same. McGee (1994) showed that if Pr is such a non-standard probability function and P a Popper function, $Pr(B/A) = P(B/A)$ for all A and B, if $Pr(A) \neq \emptyset$.

Consistency and coherence can be proved. Leblanc (1983) has given epfs for the predicate logical case and proved that the semantics stated by the above conditions plus one extra for predicate logic is both sound and complete with respect to a standard axiomatisation of classical predicate logic. Pragmatic justifications have been given for epfs by Stalnaker (1970a) and Harper (1975). Stalnaker gave a Dutch book theorem on the assumption of strict coherence. The use of strict coherence seems surprising, but is made on the assumption that a belief functions is strictly coherent iff it is coherent and P(A) = 1 just for those propositions that are true in every relevant possible outcome. Stalnaker assumes that these relevant outcomes depend on what the agent accepts.

Another kind of justification for Popper functions is given by Harper (1975). He showed that the minimal revision modelled by Popper functions satisfies some intuitive conditions on minimal change of a certain belief state. Note that according to Popper, probability and entrenchment are two almost independent notions. This has motivated authors like Harper (1975), Spohn (1987), Gärdenfors (1988) and many more to model belief revision and epistemic entrenchment in a purely qualitative framework. An acceptance state is modelled by a set of possible worlds and belief revision is constrained by minimal change with respect to the propositions accepted. Let us see to what qualitative constraints Popper functions give rise to.

The propositions believed can be recovered from epfs in the following way: $\Omega(P_T) = \{B: P(B/T) = 1\}$. Suppose now that every proposition is represented by the set of possible worlds in which it is true. On this assumption it is of course easy to go from the set of propositions the agent accepts to the possible world analysis of belief: $K(P_A) = \cap\Omega(P_A)$. The most obvious advantage of epfs above standard probability functions is of course that P(A/B) is also defined when P(B) = 0. An epf represents an extended state of belief. It represents a set of hypothetical states of belief, one for each condition. The belief-state $K(P_A)$ such that P(A) = 0 will be disjoint from the current belief-state $K(P_T)$.

Harper showed that epfs obey the following conditions on minimal change:

(ai) $K(P_A) \neq \emptyset$, for some A,

(aii) If $K(P_A) = \emptyset$, then P(B/A) = 1 for all B,

(b) P_A is coherent relative to $K(P_A)$, if $K(P_A) \neq \emptyset$, and

(c) If $K(P_A) \cap B \neq \emptyset$, then $K(P_{A \wedge B}) = K(P_A) \cap B$.

Also Gärdenfors (1988) gave conditions of minimal change which any revision function should obey. It is interesting to know that if propositions are represented by sets of possible worlds, every compact epf P gives rise to a revision function that satisfies his well known postulates (K^*1)-(K^*8) for revision.[297] I formulate these postulates on the assumption that belief states and propositions are represented by sets of possible worlds, and that K^*_A is the set of possible worlds that result from revising K by A:[298]

(K^*1) For any proposition A and any belief state K, K^*_A is a belief set.

(K^*2) $K^*_A \subseteq A$

(K^*3) $K \cap A \subseteq K^*_A$

(K^*4) If $K \cap A \neq \emptyset$, then $K^*_A \subseteq K \cap A$

---

[297] See his §5.8 for more discussion of the relation between Popper functions and the constraints given for qualitative revision.

[298] Just as the constraints on selection functions do not determine $C_K(A)$, the postulates below do not determine K^*_A uniquely.

(K^*5) $K^*_A = \emptyset$ iff $A = \emptyset$ (on the assumption that $K \neq \emptyset$)

(K^*6) If $A = B$, then $K^*_A = K^*_B$[299]

(K^*7) $(K^*_A) \cap B \subseteq K^*_{A \wedge B}$

(K^*8) If $K^*_A \cap B \neq \emptyset$, then $K^*_{A \wedge B} \subseteq (K^*_A) \cap B$[300]

That epfs satisfy Harper's constraint (c) is important to see (it takes care of (K^*7) and (K^*8) and thus also of (K^*3) and K^*4)), and this constraint will stay important below.

Stalnaker and Lewis used in their possible world analysis of conditionals a similarity function defined on single possible worlds. We have seen that in terms of these primitive similarity functions we could define a revision function on global belief states. Revisions of conditional probability functions are based on an essentially different idea. The revision is primitivily defined in terms of the *global* belief state, represented by a conditional probability function. Let us call revision functions primitivily defined in terms of global belief states *epistemic revision*. Let us now turn to another possible way of handling epistemic revision. Grove (1986) has developed a system of sphere models that represents revision functions that behaves similar to the Bayesian account of revision handled by extended probability or Popper functions.[301] He uses Lewis's system of a sphere model, but doesn't take a single world as the centre of the sphere, but a whole belief state K, the set of worlds consistent with what is believed. Just like in Lewis (1973) this sphere model can be induced by an ordering relation on possible worlds. The only difference between the ordering relations is that contrary to Lewis, Grove does not assume strong centering. Instead, Grove assumes weak centering, $\forall w'$: $w \neq w' \Rightarrow w \leq_w w'$. Let us call the smallest sphere in S that has a non-empty intersection with A, S_A. If A is $\emptyset$ we set S_A to be W. The *revision of K by A* is defined as $K^*_A =_{df} A \cap S_A$. Grove showed that if a revision function is represented by means of his system of spheres, K^*_A satisfies all the Gärdenfors postulates for revision. In terms of this way of handling revision, we can now analyse conditionals in a way similar to Stalnaker's (1970a) reading of the Ramsey test (epistemic revision):

A > C is accepted in K iff C is true in every world in K^*_A.

Note that this analysis would only validate the CEM principle for counterfactuals if every S_j in S ($j \neq 0$) consists of only one more possible world than its foregoing sphere S_i.

Remember that the logics of induction and confirmation were developed to capture the inductive (causal) relations between propositions. But that is exactly what the Stalnaker/Lewis logic of counterfactuals intended to capture, too. It is only natural that the following question sooner or later should arise: Are these ways of handling conditional beliefs, belief change, and inductive logic's related, and if so, how? The answer to those

---

[299] This assumption is of course trivial once we assume that belief states and propositions are represented by sets of possible worlds.

[300] (K^*7) and (K^*8) together are equivalent to (K^*7,8), If $K^*_A \cap B \neq \emptyset$, then $K^*_{A \wedge B} = (K^*_A) \cap B$. It is useful to know that (K^*7) and (K^*8) together imply the following criterion: $K^*_A \subseteq B$ and $K^*_B \subseteq A$ iff $K^*_A = K^*_B$, a stronger version of a criterion already familiar to us. Finally, note that (K^*3) and (K^*4) are special cases of (K^*7) and (K^*8).

[301] Other influential representations of revision functions are given by Kratzer (1981), Alchourron, Gärdenfors and Makinson (1985) and Spohn (1987). Kratzer, in distinction with the other two, gives a *distributive* analysis of revision and thus stays closer to Stalnaker (1968) and Lewis (1973) in at least one sense of closeness. The revision method of Spohn can be thought of as a qualitative version of Jeffrey conditionalisation.

questions would clarify something about how a set of possible worlds partly determines the selection function.

Stalnaker (1970a) made the following strong but also very natural proposal: the probability of truth of a conditional equals the conditional probability. This proposal does not only mean that conditionals in general could equally well be analysed epistemically in the Bayesian tradition using epfs. But also that the two different analyses of revision (the *distributive* one of Stalnaker/Lewis and the *global* epistemic one) come down to the same. Assuming that the minimal revision of a belief state in terms of a similarity function is equal to the minimal revision of a belief state in terms of conditionalisation, his natural proposal was that $P(A > C) = P(C/A)$.[302] This proposal is known as *Stalnaker's hypothesis*. To be a bit more precise, let <W, F, P> be a probability space where W is a set of worlds, F a field of subsets of W closed under the Boolean operations, and P an arbitrary probability function closed under conditionalisation. Stalnaker implicitly assumed that ">" has a fixed interpretation. His hypothesis was that there is a binary connective ">" that behaves like a conditional such that for any probability space <W, F, P>, such that for all $A,B \in F$, $A > B \in F$, and where the probability function can represent a rational agent's system of belief, it is the case that for all $A,B \in F$: $P(B/A) = P(A > B)$, if $P(A) > 0$.[303] Because P is closed under conditionalisation, if P_C is the probability function that results from P by conditionalising on C, the hypothesis also says that $P_C(B/A) = P_C(A > B)$, if $P_C(A) > 0$. Stalnaker's main interest in this hypothesis was that, if this were true, it would give an independent argument in favour of his controversial conditional excluded middle (CEM) principle.[304]

Stalnaker's hypothesis can also be stated in a qualitative way. The hypothesis then says that there is a binary connective ">" that behaves like a conditional such that for any K, $A > B$ is accepted in K iff B is accepted in K^*_A. This in turn comes down to the hypothesis that there is a selection function f such that for any K and A, $K^*_A = \cup\{f_W(A) | w \in K\}$.[305]

5.5 Triviality

Stalnaker (1968) assumed that the correct account of conditionals should be based on the Ramsey test analysis, $A > C$ is accepted in K iff B is accepted in K revised by A. In Stalnaker (1970a) it is assumed that revision should be handled as in the epistemic Bayessian approach. These two assumptions together gave rise to the hypothesis, $P(A > C) = P(C/A)$. Lewis's triviality result showed, however, that this hypothesis is false for all but some trivial probability functions.[306] More in detail, Lewis showed that any probability function that satisfies Stalnaker's hypothesis and the constraint that the probability function is iterative:

(CSH) $\quad P(A > B/C) = P(B/A \wedge C)$, if $P(A \wedge C) \neq 0$

---

[302] See Hajek & Hall (1994) for a discussion of related hypotheses by Adams and others.

[303] Stalnaker does not really demand that $P(A) > 0$, but that is not important here.

[304] Because for proposition A that has a non-zero probability, by definition $P(\sim B/A) = 1 - P(B/A)$. Assuming that Stalnaker's proposal were true, both $P(\sim(A > B))$ and $P(A > \sim B)$ would have the same value as $P(\sim B/A)$. From this we could then immediately derive Stalnaker's CEM.

[305] In Gärdenfors (1988), <**K**, *> is called a belief revision model, where **K** is a set of belief sets and * a revision function. Gärdenfors assumes that $A > B \in K$ iff $B \in K^*_A$ and that for any $K \in \mathbf{K}$ and any proposition A, the revision of K by A, K^*_A, is again an element of **K**. In other words, what is assumed is that there is a * such that any $K \in \mathbf{K}$ and any proposition A, the revision of K by A, K^*_A, is again an element of **K**, and that $A > B \in K$ iff $B \in K^*_A$.

[306] For a clear exposition of Lewis's proof and some much more telling results, see Hajek & Hall (1994).

for any binary connective >, can only assign different probabilities to two different propositions.

As we will see, the extra constraint (CSH) follows from Stalnaker's hypothesis extended to conditional probabilities, and by the assumption that conditional probabilities satisfy the standard laws.

According to any standard analysis of probability, the result of successive conditionalisation on two statements is the same as that of conditionalising once on the conjunction of those statements. This can be illustrated by the following example:

Suppose an unbiased coin is tossed two times. The value our subjective probability function P will assign to head of the second toss, P(h2), will be 1/2. After we learn that at least one of the two tosses yielded head, our probability assigned to h_2 will be $P(h_2/h1 \vee h_2) = (1/2)/(3/4) = 2/3$. Let's call the new resulting probability function P'. If we learn that the two tosses did not both yield head, we conditionalise P' by $\sim(h_1 \wedge h_2)$, $P'(h_2/\sim(h_1 \wedge h_2)) = P'(h_2 \wedge \sim(h_1 \wedge h_2))/P'(\sim(h_1 \wedge h_2)) = (1/3)/(2/3) = 1/2$. So the probability of h_2 according to the probability function P'' resulting after two times conditionalising is 1/2. The same results if we conditionalise once on the conjunction of those two statements, $P''(h_2) = P(h_2/(h_1 \vee h_2) \,\&\, \sim(h_1 \wedge h_2))$.

Suppose now that (B/A)/C makes sense as a statement to which a probability function P can be applied that obeys the usual conditions. That is, let us make the crucial assumption of Stalnaker (1970a), i.e. that '/' obeys conditionalisation *and* is a connective with a context independent meaning. Then it follows that for any probability function P, $P((B/A)/C) = P(B/A \wedge C)$, if $P(A \wedge C) \neq 0$. This then, really is (CSH). The condition (CSH) can be proven on the following definition of a subfunction, and the four assumptions below:

Definition:
A *subfunction*, P_A, is a function defined for any probability function P and proposition A such that $P(A) \neq 0$ as follows: $P_A(B) =_{df} P(B/A)$

Assumptions:

(0) P(B/A) is defined only when $P(A) \neq 0$,
(1) If $P(A) \neq 0$, $P(A > B) = P(B/A)$, Stalnaker's hypothesis,
(2) Any subfunction is a probability function,
(3) The conditional has a fixed interpretation, "B > C" expresses the same proposition in all contexts/probability functions.

The assumption that the conditional has a fixed interpretation is used by Lewis in the following form: ">" means the same in P and in P_C, for any C and P.

On the basis of these assumptions, we can prove (CSH):

(CSH) $P(A > B/C) = P(B/A \wedge C)$, if $P(A \wedge C) \neq 0$.

Lewis (1975a) derived the triviality result from (CSH), which follows from the assumptions made in Stalnaker (1970a). Stalnaker (1976b) showed how (CSH) follows from his assumptions:

$P_C(A \wedge B) = P_C(A) \times P_C(B/A)$ (by axioms of P)
$= P_C(A) \times P_C(A > B)$, if $P_C(A) \neq 0$, (by (1))

(a)	=	$P(A/C) \times P(A > B/C)$, if $P(A \wedge C) \neq 0$,	(by subfunction and (3))
$P_C(A \wedge B)$	=	$P(A \wedge B/C)$	(by subfunction)
(b)	=	$P(A/C) \times P(B/A \wedge C)$	(by axioms of P)
If $P(A \wedge C) \neq 0$, $P(A > B/C) = P(B/A \wedge C)$			(by (a) and (b))

For Lewis's triviality proof, we first prove an independence property, saying that $P(A \wedge B) = P(A) \times P(B)$. The proof of this independence property is based on (CSH) and the following two standard assumptions:

(4) $P(A) = P(A/B) \times P(B) + P(A/\sim B) \times P(\sim B)$ if $0 \neq P(B) \neq 1$, expansion by cases

(5) if $P(B) \neq 0$, then (a) if $A \models B$, then $P(B/A) = 1$, and (b) if $A \models \sim B$, then $P(B/A) = 0$.

The essential step to prove that $P(C \wedge A) = P(C) \times P(A)$, is to show that on assuming (CSH) one can derive that $P(C/A) = P(C)$:

$P(C/A)$ = (4)	$P(A > C/C) \times P(C) + P(A > C/\sim C) \times P(\sim C)$, if $P(C) \neq 0 \neq P(\sim C)$	
= (CSH)	$P(C/C \wedge A) \times P(C) + P(C/\sim C \wedge A) \times P(\sim C)$, if $P(C \wedge A) \neq 0 \neq P(\sim C \wedge A)$	
= (5)	$1 \times P(C)$	$+ 0 \times P(\sim C)$
=	$P(C)$	

By conditionalisation: $P(C \wedge A) = P(C/A) \times P(A) = P(C) \times P(A)$

So, from Stalnaker's hypothesis together with the assumption that the conditional has a fixed interpretation, it follows that $P(A \wedge (A > C)) = P(A) \times P(A > C)$. It is thus predicted that $P(A)$ and $P(A > C)$ are probabilistically independent of each other.

Assuming $P(A \wedge B) = P(A) \wedge P(B)$ for any A and B, we can prove Lewis's triviality result:

If $P(A) \neq 0$, $P(A > B) = P(B/A)$, any probability function P that uses conditionalisation can assign to at most two pairwise incompatible propositions a non-zero probability.

Proof:
Let C, D and E three pairwise incompatible propositions with non-zero probability. Assume $A = C \vee D$ and Stalnaker's hypothesis. By the incompatibility of C and D it follows that $P(A \wedge C) = P(C)$, and by Stalnaker's hypothesis it follows that $P(A \wedge C) = P(A) \times P(C)$, because it is predicted that A and C are probabilistically independent. As a result it is predicted that $P(A) \times P(C) = P(C)$. Thus, $P(A) = 1$ and $P(\sim A) = 0$. But this is impossible because $P(\sim A) \geq P(E)$, which has by hypothesis a non-zero probability.

So even without assuming that $A > C$ obeys Stalnaker's logic C2, Stalnaker's hypothesis, $P(A > B) = P(B/A)$, cannot be made.

Over the years, a number of authors have strengthened and generalised Lewis's triviality proof for both probabilistic representations of information states and qualitative variants

thereof.[307] The qualitative version of Stalnaker's hypothesis says that conditionals state context independent propositions and should be handled by the Ramsey test analysis. The triviality proof (see Gärdenfors, 1988) is then based on the assumption that revision should be very much like conditionalisation in that it has to satisfy the following constraint:

(K*7,8) If $K^*_A \cap B \neq \emptyset$, then $K^*_{A \wedge B} = (K^*_A) \cap B$.

Those two assumptions together lead to the qualitative version of (CSH) (where $K^*_A \models B$ means that B is accepted in K revised by A):

(CSH) $K^*_C \models A > B$ iff $K^*_C \cap A \models B$, if $K^*_C \cap A \neq \emptyset$

Note that the following constraint is a special case of (K*7,8):

(K*3,4) If $K \cap A \neq \emptyset$, then $K^*_A = K \cap A$.

If a revision function satisfies (K*3,4), the revision function is called *preservative*.

What the triviality results at least show is that the following four conditions are not jointly satisfiable:

(a) conditionals should be analysed via the Ramsey test,[308]
(b) all conditionals state propositions,
(c) the conditional has a fixed interpretation,[309] and
(d) the revision function satisfies (CSH) or its qualitative variant (K*7,8).

Note that the third condition is the basic assumption behind the original similarity account of counterfactuals, while the fourth condition is assumed in all global revision methods.

5.6 Reactions to triviality

Given that the four above principles are jointly responsible for the triviality results, it's clear that we can react in at least four ways to the results of Lewis and others. And indeed, this is what happened. The present section gives a survey of these attempts.

5.6.1 Imaging versus epistemic revision

After destroying Stalnaker's hypothesis, Lewis (1975a) showed that we can keep the Ramsey test analysis for conditionals in general by giving up (CSH) and defining revision in terms of imaging. But what is imaging?

Imaging is a function of minimal belief change which uses not primarily the information available in the information state ordered by epistemic entrenchment (as in Stalnaker, 1970a, and Grove), but the similarity relation between *individual* possible worlds. The

---

[307] See Gärdenfors (1988) for a qualitative variant of the triviality proof, and Hajek & Hall (1994) for an overview. Gärdenfors's impossibility proof is a strenghtening of Lewis's triviality proof, because there is no qualitative variant of the expansion by cases rule of probability functions.

[308] $A > C$ is accepted in K iff C is accepted in K revised by A, where A and C can be any proposition, and where A is accepted in K means that A is *true* in all possibilities of K.

[309] If the conditional is analysed via the Ramsey test and the conditional has a fixed interpretation, the following monotonicity principle follows immediately: $K \subseteq K'$ & $K'^*_A \subseteq B \Rightarrow K^*_A \subseteq B$. Proof (see also Gärdenfors, 1988, p. 157): Let us assume that $K'^*_A \subseteq B$, then by the Ramsey test $K' \subseteq A > B$. Because $K \subseteq K'$ it follows that $K \subseteq A > B$, and by the Ramsey test again $K^*_A \subseteq B$.

consequence is that it differs from conditionalisation (of normal or conditional probability-functions) in an interesting way. Here is the intuition behind it:

> Imaging P on A gives a minimal revision in this sense: unlike all other revisions of P to make A certain, it involves no gratuitous movement of probability from worlds to dissimilar worlds. Conditionalisation P on A gives a minimal revision in this different sense: unlike all other revisions of P to make A certain, it does not distort the profile of probability ratios, equalities, and inequalities among sentences. (Lewis, 1975a)

Let us think of the probability functions as assigning probabilities to the (finitely many) worlds such that the probabilities add up to one. Let us now assume (by the uniqueness assumption) that for every world *w* and proposition A there is a unique world $f_w(A)$. Given a probability function P and any possible A, there is a probability function P_A such that, for any world *w'*:

$$P_A(w') = \Sigma_w P(w) \times \begin{cases} 1 \text{ if } f_w(A) \text{ is } w' \\ 0 \text{ otherwise} \end{cases}$$

Lewis calls P_A the image of P on A, and says that P_A comes from P by imaging on A. Intuitively, the image on A of a probability function is formed by shifting the original probability of each world w over to $f_w(A)$. Then Lewis is able to prove (unsurprisingly) that $P(A > C) = P_A(C)$.[310]

What is interesting about imaging is that the preservation property for revision is no longer valid. Think of K as the information state before revision defined in the following way: $K = \cap\{B: P(B) = 1\}$, where each proposition B is represented by the set of possible worlds in which it is true. Revising the information state K by a proposition A that is consistent with it does not result necessarily in an information state K' that is a subset of K. The reason is that even if $K \cap A \neq \emptyset$, the most similar A-world to a *w* in K that doesn't make A true, doesn't have to be an element of K. For this reason it might falsify a proposition that was verified by every element of K. Note that the state $K' = \cap\{B: P_A(B) = 1\}$ can also be determined by our earlier change function C, $K' = C_K(A)$. We can conclude that also the qualitative revision function C is not in general preservative, i.e. it does not generally satisfy (K^*7,8), the qualitative version of (CSH). Now we can understand why Stalnaker's (1970a) proposal was not justified, his assumption that the minimal revision of a belief state in terms of a context independent similarity function is equal to the minimal revision of a belief state in terms of conditionalisation was wrong, K^*_A need not be the same as $C_K(A)$.

Lewis (1975a) showed that we can keep the Ramsey test analysis for counterfactuals, if we give up (CSH). Van Fraassen (1976) even showed that if we give up (CSH) a version of Stalnaker's hypothesis might still be true.

5.6.2 Van Fraassen

Discussing the triviality result, we have seen that any probability function that satisfies Stalnaker's hypothesis and principle (CSH), $P(A \Rightarrow B/C) = P(B/A \wedge C)$, if $P(A \wedge C) \neq 0$, for any binary connective $\Rightarrow$, will be trivial. We have also seen that one of the premises for

---

[310] Where $P_A(C) = \Sigma_w P(w) \times \begin{cases} 1 \text{ if } f_w(A) \in C \\ 0 \text{ otherwise} \end{cases}$

Stalnaker didn't really show much interest in this last result. The reason should be obvious. Even though it also verifies the principle CEM, it can hardly be called independent motivation for it. And indeed, Gärdenfors (1982) showed that imaging can also be defined without the uniqueness assumption. The result is that the probability originally assigned to a world where A is not true is possibly spread over more than one world where A is true. This obviously reflects Lewis's analysis of counterfactuals, instead of Stalnaker's, in that it doesn't validate CEM anymore.

deriving (CSH) was the assumption that $\Rightarrow$ has a fixed interpretation. Van Fraassen (1976) called this assumption *metaphysical realism* and proposed to give that up. Giving up the assumption that conditionals have a fixed interpretation, Van Fraassen was able to prove that for every probability function there is a binary connective ">" such that it has the same meaning in both occurrences of the embedded conditional "(A > B) > C", and where both P(A > B) = P(B/A), and CEM holds.[311] Note that Van Fraassen's result is much weaker than that what first was proposed by Stalnaker. Stalnaker's (implicit) hypothesis was that there is a ">" such that for all P, P(A > C) = P(C/A), whereas Van Fraassen only proved that for every P there is a ">" such that P(A > C) = P(C/A). Making ">" context dependent is compatible with Van Fraassen's claim, but not with Stalnaker's hypothesis as originally intended. Van Fraassen's result is weaker than the original hypothesis in another respect, too. As shown in Stalnaker (1976b), the probability of A > C cannot be equal to P(C/A) such that ">" obeys Stalnaker's logic, even if ">" is made context dependent. Hajek & Hall (1994) showed that ">" cannot even obey Lewis's logic. Indeed, in Van Fraassen's logic CE, the axiom that corresponds with the following constraint on selection functions is given up, if $f_W(A) \subseteq B$ and $f_W(B) \subseteq A$, then $f_W(A) = f_W(B)$, a constraint shared by Stalnaker and Lewis.[312]

5.6.3 Two kinds of belief change

If conditionalisation and expansion are special kinds of revision, we might say that Van Fraassen proposed to give up the assumption that revision should obey (K^*7,8), or its corresponding probabilistic version. But this seems to me a very unnatural reaction, because this constraint seems to capture exactly what is going on if we change our beliefs by learning new information. Moreover, the principle also seems to be needed to account for (embedded) indicative conditionals. Giving up preservativity or the assumption from which it follows doesn't seem to be the right way to go. Indicative conditionals are only appropriately asserted in a given context if their antecedents are consistent with the context, if a context is represented by a set of worlds. To interpret the consequent, we should only consider worlds in the context

Contrary to Van Fraassen, Stalnaker responded to the triviality result of Lewis by giving up his hypothesis and by arguing that, on second thought, the probability of truth of a counterfactual *should* not be equal to its corresponding conditional probability. Remember that the Bayesian account of probability is purely epistemic in nature. So P(C/A) > P(C) means that A is *evidentially* relevant for the acceptance of C. But if his original analysis of counterfactuals is an appropriate analysis of causal relations and if Stalnaker's proposal were true, evidential relevance would be equal to *causal relevance*. But this is clearly not true and some puzzles in Jeffrey's (1965) purely evidential decision theory made this clear. According to Jeffrey's decision theory, actions are evaluated according to the probability the deliberator assigns to the desired state conditional on the proposition expressed by the action. The conditional probability P(C/A) models the evidential relation the agent sees between A and C; if P(C/A) is high, the agent would assign a high probability to C, if he would *learn the news* that A is the case. Obviously, if A causes C, P(C/A) would be high, but the problem is that P(C/A) might also be high in cases where A does not cause C, but where both are caused by a common cause. Stalnaker (1980b) gave the following example: Suppose that the correlation between smoking and long cancer was not due to the consequences of smoking for the lungs, but due to a common genetic factor that causes both the tendency to smoke and the tendency to develop lung cancer. In that case there is no reason for agents to withdraw smoking in order to prevent lung cancer, although the

---

311 See also Gibbard (1980).

312 For more discussion see Jeffrey & Stalnaker (1994), and for reasons to be suspicious, see Hajek & Hall (1994). It should be noted that Van Fraassen also had a second method of saving Stalnaker's hypothesis in the $\forall\exists$ form for triviality, viz. by restricting the hypothesis to a limited class of conditionals. In that case, the logic for conditionals is allowed be as strong as Stalnaker's logic.

probability of getting a lung cancer conditional on smoking is high. Stalnaker concluded that causal relevance, the kind of relevance needed to evaluate one's actions in a deliberation, should not be modelled by conditional probabilities of consequences with respect to actions. He suggested that, instead, the use of conditional probabilities in Jeffrey's theory should be replaced by the probabilities of their counterfactuals expressed.[313] This suggestion has been worked out by various authors and resulted in *causal decision theory* (see Gibbard & Harper (1978)).

With the distinction between evidential and causal decision theory, there corresponds a distinction between two ways of changing one's belief state. Conditionalisation is supposed to mirror the way a rational agent would change his belief state if he would learn new information, while imaging is supposed to mirror the way a rational agent would change his belief state if he, or somebody else, would do a certain action. The distinction between conditionalisation and imaging has in turn its qualitative correspondence; in Katsuno & Mendelzon (1991) the qualitative version of conditionalisation is called the *revision* of a belief state, and by the *update* of a belief state is meant the qualitative version of imaging.[314]

Stalnaker argued that P(A > C) and P(C/A) should in general not be the same. Global revision, and distributive revision by imaging reflect a different intuition. A number of people have suggested that this difference corresponds with a difference between indicative conditionals and counterfactuals. According to this suggestion, an assertion of an indicative conditional mirrors the conditional probability the speaker assigns to the consequent with respect to the antecedent, while what is expressed by a counterfactual is less directly dependent on the speaker's current belief state. In the following sections we will discuss various ways in which these suggestions have been implemented.

5.6.4 Adams

According to Adams (1970) there exists a difference between indicative and subjunctive conditionals. He motivated this distinction by noting that if *Oswald* is in focus there is a difference between accepting (4) and (5):

(4) If Oswald didn't shoot Kennedy then someone else did.
(5) If Oswald hadn't shot Kennedy, someone else would have.

If we learn that Oswald did not shoot Kennedy, we would immediately accept that somebody else did, but it is not so clear that we would accept that someone else would have killed Kennedy if Oswald hadn't shot him. Adams proposed already in the sixties that P(B/A) should not be equated with the probability of *truth* of A > B, but rather with its *assertability*. This is not only the case for indicative, but also for counterfactuals conditionals. The difference between (4) and (5) is then explained by choosing a different probability function[315] for the analysis of an indicative conditional and its corresponding counterfactual. Because P(B/A) is not equated with the probability of truth of A > B, Adams need not assume that A > B states a proposition. It can be argued that what the triviality result really showed is that conditionals do not express propositions. The problem with this suggestion is that if conditionals no longer express propositions, it is not clear anymore how to account for embedded conditionals.

---

[313] In the non-backtracking reading.

[314] Although it is questionable whether the Lewis/Stalnaker constraints on selection functions, or the postulates for updating of Katsuno & Mendelzon exactly capture the distinct features of causal dependence.

[315] He proposed in Adams (1976) that for counterfactuals, not the current, but a *prior* probability function is relevant. For a somewhat different proposal, see Skyrms (1994).

5.6.5 Lewis

Also for Lewis (1975a) the distinction between indicative and subjunctive conditionals is a real one. He was happy to give up (in fact, never defended) the assumption that we should analyse all conditionals in a uniform way by the Ramsey test analysis.[316] Lewis never accepted the global revision approach for subjunctive conditionals. For his way of handling counterfactuals, and the probability thereof, the triviality result was not disturbing. In contrary, the triviality proof showed that the independent motivation for principle CEM, that he had argued against before, was of no good. Lewis also did not agree with Stalnaker that indicative conditionals should be handled in the same way as counterfactuals. According to Lewis the difference between the two corresponds with a *semantic* distinction. But he did not agree with Adams that indicative conditionals do not express propositions. Indicative conditionals state propositions, but should be analysed in terms of the material implication. To account for the paradoxes of the material implication he proposed to rely on Grice. He was prepared to admit to Adams that the assertability of indicative conditionals goes by conditional probability.[317] But claiming that the truth conditional content should be handled by the material implication enables him to account for iterated conditionals.

Lewis's analysis of indicative conditionals is, however, not very natural, because it treats the antecedent and the consequent of indicative conditionals symmetrical with respect to truth, but asymmetrical with respect to assertability. But this is problematic if there are examples where a conditional has intuitively a different assertability value as another sentence that by the material implication account of conditionals is truth functionally equivalent with it. Examples of this kind have been given by various authors, but the proponent of the material implication account can always argue that the difference is not due to truth conditional content, but to the *form* in which this content is asserted. But even this defence strategy doesn't work anymore once we embed two such clauses into a bigger sentence that intuitively have a different assertability or even truth value. Gibbard (1980) has given such an example, but maybe the greatest difficulty for a Gricean account towards indicative conditionals is given by Grice (1967) himself. Consider the case where Yog and Zog play chess, Yog has white 9 out of 10 times, and draws are not allowed. We don't know who won what game, but we do know that of the hundred games they played up to now, Yog won 80 times when he had white and lost all 10 times that he had black. Intuitively, the following two assertions are true of any one of the hundred arbitrary games they played:

(6) If Yog had white, there is a probability of 8/9 that he won.
(7) If Yog didn't win, there is a probability of 1/2 that he didn't have white.

The problem for the material implication account is that it cannot account for the truth of these assertions. If the probability operator has scope only over the consequent, we again have the well known paradox of the material implication. If the probability operator takes scope over the whole conditional, the material implication account would predict that the embedded sentences of (6) and (7) are truth functionally equivalent. But how can that be if their assigned probability is different?[318] To treat both assertability and truth in a similar way, it appears natural to use Belnap's three-valued (and two-dimensional) analysis of

---

316 Also Gärdenfors (1988) responded to the triviality result by giving up the assumption that all conditionals should be analysed via the Ramsey test.

317 Note that if indicative conditionals are analysed by material implication, probability of truth and conditional probability equal each other only in extreme cases. On the other hand, Lewis (1975a) showed that the conditional probability equals the probability of the material implication minus the probability of those cases in which asserting the conditional would be misleading: $P(B/A) = P(A \rightarrow B) - [P(\neg A) \times (P(A \wedge \neg B)/P(A))]$.

318 The problem is of course that according to the material implication account, contraposition is valid. There are other examples suggesting that contraposition should not be valid for indicative conditionals: from 'If it is after 3 o'clock, it is not much after 3 o'clock' we don't infer to 'If it is much after 3 o'clock, it is not after 3 o'clock' (Nute, 1984, p. 428).

conditionals to determine the truth-value of indicative conditionals. In this way, for most indicative conditionals at least, conditional probability equals its probability of truth (see Skyrms, 1980a, p. 89).

5.6.6 The preservativity principle

Contrary to Lewis, Stalnaker claimed that the similarity analysis of conditionals can be used for both counterfactuals and indicative conditionals. However, this gives rise to a problem; the following argument is not valid:

(8) Either the butler or the gardener did it. Therefore, if the butler didn't do it, the gardener did.[319]

To account for this intuitively valid argument, Stalnaker introduced the notion of *reasonable inference*, a pragmatic relation between speech acts instead of the semantic relation of entailment between propositions. C is a reasonable inference of $A_1,...,A_n$ iff the content of C is entailed by the context resulting from the initial context updated by $A_1,...,A_n$, provided that for each $i \leq n$, the assertion A_i is made in an appropriate initial context.[320] The above direct inference is a reasonable inference if the following assumptions are made for being appropriate contexts for disjunctions and indicative conditionals:

(a) If an indicative conditional is being evaluated at a world in the context set, then the world selected must, if possible, be within the context set as well.[321]
(b) A disjunctive statement is appropriately made only in a context which allows either disjunct to be true without the other.

Suppose that $A \vee C$ is appropriate in a given context. It follows that $C \wedge \neg A$ is compatible with the context set that represents the presupposed information. If then $A \vee C$ is added to the context set, the antecedent of the conditional *if ~A, then C* will be compatible with the new context set. Because all ~A-worlds in the context set are C-worlds, and because by (b) the selected ~A world will be a world in the context set, the inference in (8) is reasonable.

Let K be a presupposition state, then we might say that for indicative conditionals with antecedent A, Stalnaker's appropriateness condition (a) has the following principle as consequence:

A-worlds in K are to be selected as nearer to worlds in K than any A-world outside of K.

This principle seems to be the only reasonable assumption to make for an appropriate analysis of indicative conditionals. The principle is known as the principle of *preservativity*. We can follow Harper (1976) and Morreau (1992) and implement this principle by relativising the selection function to the belief state. Given the definition of $f_W(A)$, we can relativise our selection function to a context K in the following way:

$$f^K_W(A) \quad = \quad f_W(A \cap K), \text{ if } w \in K \text{ and } A \cap K \neq \varnothing$$

---

[319] If the conditional is analysed as the material conditional, the argument is predicted to be valid. But Stalnaker rejects this analysis for indicative conditionals because it leads to a lot of other well known problems.

[320] See the Appendix to Stalnaker (1975) for more details.

[321] If all indicative conditionals obey (a), it gives rise to the following appropriateness condition: It is appropriate to make an indicative conditional statement or supposition only in a context which is compatible with the antecedent. For a motivation for this principle, see Stalnaker (1975). Two other inferences that are invalid according to Stalnaker's semantics for conditionals, contraposition and the hypothetical syllogism, turn out to be reasonable for indicative conditionals.

$= \quad f_w(A)$, otherwise.[322]

Now we can determine what proposition is expressed by the indicative conditional *If A, then B* in context K, $A >_K B$:

$A >_K B \quad = \quad \{w \in W: f^K{}_w(A) \subseteq B\}.$

In terms of the relativised selection function, we can also define the following context dependent revision function, $C'_K(A)$, the revision from K with A:

$C'_K(A) \quad = \quad \{f^K{}_w(A): w \in K\}$

As especially made clear in Morreau (1992), when we accept the preservativity principle for selection functions, when the belief state changes, the selection function changes too. Let **K** be a belief state represented by <K, f>, where **K** is a set of worlds and *f* a selection function. When we revise **K** by A, the new belief state will be of the following kind: $<\cup\{f^K{}_w(A): w \in K\}, g>$, where g is a selection function, a function from worlds and propositions to propositions, that satisfies the following conditions for all propositions A:

(a) $g_w(A) \subseteq A$,

(b) $w \in A \Rightarrow w \in g_w(A)$, and

(c) g obeys the preservativity principle with respect to $\cup\{f^K{}_w(A): w \in K\}$[323]

Morreau (1992, §2.6) showed that if the preservativity principle is assumed this extra dynamic element of belief change can handle examples (Hansson's example, and Tichy's example) that are problematic if it is assumed that the selection function doesn't change. Let us here only discuss Hansson's example in the abstract. Suppose our language contains only two propositions, A and B. In the first story it is only presupposed that $A \vee B$ is true, and thus $\sim A > B$. If the information that A is added to the information state, the information state will still accept the propositions $A \vee B$ and $\sim A > B$. In the second story, A is added directly to the information state. From this it follows that also $A \vee B$ is accepted, but intuitively there is no reason why we should accept $\sim A > B$. we cannot rule out that $\sim A > \sim B$ is true. The difference between the two story's cannot be explained if a fixed selection function is assumed. Let us consider the following belief state <K,f> to begin with:

$W = K = \{w_1, w_2, w_3, w_4\}$, $A = \{w_1, w_2\}$, $B = \{w_1, w_3\}$ and $f_{w1}(\sim A) = f_{w2}(\sim A) = \{w_3, w_4\}$ (the other similarity relations are irrelevant)

By learning $A \vee B$, <K, f> changes into <K',g>. The selection function g is different from f because it has to obey preservativity. <K',g> will verify $\sim A > B$.

$W = \{w_1, w_2, w_3, w_4\}$, $K' = \{w_1, w_2, w_3\}$, $A = \{w1, w_2\}$, $B = \{w_1, w3\}$ and $g_{w1}(\sim A) = g_{w2}(\sim A) = \{w_3\}$

If we now learn A, the final information state will be <K",h>, which still verifies $\sim A > B$.

---

[322] The ordering relation in terms of which the selection function is defined still obeys centering, transitivity, and connectedness.

[323] Morreau also demands that the new selection function g must be maximal on the partial order $\leq$ defined as follows: $f \leq g$ iff $\forall w \in W, \forall A \subseteq W: f_w(A) \subseteq g_w(A) \; \& \; f_w(A) \neq g_w(A)$

$W = \{w_1, w_2, w_3, w_4\}$, $K'' = \{w_1, w_2\}$, $A = \{w_1, w_2\}$, $B = \{w_1, w_3\}$ and $h_{W1}(\sim A) = h_{W2}(\sim A) = \{w_3\}$

However, if we go from <K, f> directly to a belief state where A is true, the information state that result is <K''', j> that does not verify ~A > B:

$W = K = \{w_1, w_2, w_3, w_4\}$, $A = \{w_1, w_2\}$, $B = \{w_1, w_3\}$ and $j_{W1}(\sim A) = j_{W2}(\sim A) = \{w_3, w_4\}$

Tichy (1976) has given an example showing that similarity cannot be measured by all propositions that are accepted in a belief state. That would have as a result that if A does not entail ~B, and B is accepted in K, then B is also accepted in K revised by A. But now suppose that we believe of a man that he always wears his hat when it is raining, but that it is completely impredictable whether he wears his hat when it is not raining. Suppose now that we learn that it is in fact raining. If we assumed that similarity was measured by all accepted propositions, we would predict that we also come to believe that the following counterfactual would be true: *If it had not rained, the man would also wear his hat*, although intuitively there is no reason why this should be so. As shown by Morreau (1992), if we only accept the preservativity principle, Tichy's example can be easily accounted for. Let our initial belief state be <K,f>, where $K = W = \{w_1, w_2, w_3, w_4\}$, R (it is raining) = $\{w_1, w_2\}$, WH (wears hat) = $\{w_1, w_2, w_3\}$, $f_{W3}(R) = f_{W4}(R) = R$, and $f_{W1}(\sim R) = f_{W2}(\sim R) = \sim R$. After we learn that it rains, our new belief state will be <K',g>, where $K' = K \cap R$, and *g* is the same as *f* with respect to the propositions R and ~R. In that case the conditional ~R > WH is still not accepted in <K',g>.[324]

Accepting the appropriateness condition for indicative conditionals has the effect that the meaning of the conditional connective becomes context dependent. It depends on what is commonly presupposed by participants of a conversation. However, this proposal is problematic. First, it is not clear how to account for Grice's example.[325] Second, Gibbard has given an example to which we will now turn, showing that accepting the preservativity principle won't be enough to account for all cases of indicative conditionals.

5.6.7 Gibbard

A number of authors have observed that the interpretation or assertability of an indicative conditional is much more context dependent than that of a subjunctive one. To account for this difference, Gibbard (1980) argued that while subjunctive conditionals can express context independent propositions and should be analysed in terms of Lewis's and Stalnaker's original similarity account, indicative conditionals are more closely related to the epistemic state of the agents who utter them, and should be analysed via the Ramsey test analysis. The latter suggestion can be implemented in two ways. Either we follow Adams and Belnap and claim that by uttering indicative conditionals we do not always express propositions, but instead make conditional assertions. What is asserted then depends on what is believed by the speaker. The other possibility is that we still demand that indicative conditionals always express propositions, that those conditionals are handled via the Ramsey test, but that we give up the assumption that the conditional has a fixed interpretation. We have seen that Stalnaker suggested something like the latter approach. Gibbard argues that the first approach is to be preferred, because contrary to Stalnaker's analysis it can account for the paradoxical fact that people who believe the conditional *if A,*

---

[324] That we can account for Tichy's example by assuming only the first principle of Harper (1976) does not mean that we cannot account for the example if we also assume his second principle. We could if we assumed that in this case the accepted proposition WH does not determine similarity.

[325] Although ">" does not validate contraposition, we know that P(B/A) is in general not the same as P(A > B). Sometimes they cannot even be the same, as shown by the following example: $W = \{w,w',w''\}$, $w \in A \cap B$, $w' \in A \cap \neg B$ and $w'' \notin A$. If all worlds have equal probability, P(B/A) = 1/2, but P(A > B) is either 2/3, if $f_{w''}(A) = w$, or 1/3, if $f_{w''}(A) = w'$.

then B can come to accept the opposite conditional *if A, then not B* and learn something from it, without having to revise their old belief state.

One of the central features of Stalnaker's (and Lewis's) conditional logic is the *principle of conditional non-contradiction*, the assumption that A > B is inconsistent with A > ~B. This in distinction with the material implication: out of A → B and A → ~B, you cannot derive a contradiction but instead conclude ~A. Gibbard (1980) has given a very nasty example that shows a problematic aspect of the principle of conditional non-contradiction:

> Sly Pete and Mr. Stone are playing poker on a Mississippi riverboat. It is now up to Pete to call or fold. My henchman Zack sees Stone's hand, which is quite good, and signals its content to Pete. My henchman Jack sees both hands, and sees that Pete's hand is rather low, so that Stone's is the winning hand. At this point the room is cleared. A few minutes later Zack slips me a note which says "if Pete called, he won", and Jack slips me a note which says "if Pete called, he lost". I know that these notes both come from my henchmen, but do not know which of them sent which note. I conclude that Pete folded. (Gibbard, 1980)

Gibbard argues that if both utterances express propositions, both of them should be accepted as true. But this is inconsistent with the principle of conditional non-contradiction. Because Gibbard does not believe that conditionals should be handled by the material implication, he concludes that indicative conditionals do not express propositions. Instead, they are conditional assertions that mimic the probability the speaker assigns to the consequent conditional on the antecedent. He also argues that this non-propositional account of indicative conditionals has an extra advantage, it explains why many embeddings of indicative conditionals doesn't seem to make sense. Embeddings to the right, A > (B > C) are not so problematic for the probabilistic account, if it is assumed that they are equivalent to (A ∧ B) > C.

5.6.8 A unified account

Assuming that indicative conditionals do not state propositions is problematic, however. Indicative conditional embedded to the left are then difficult to handle, although (at least sometimes) they do make sense.

(9) If the cup broke if dropped, then it was fragile. (Gibbard, 1980)

Moreover, it doesn't seem very plausible to assume that indicative conditionals should be analysed so differently from subjunctive conditionals. It is unwanted because it cannot be explained anymore why both kind of conditionals use the same words, combine with the same functions (even if, only if , if..might) in similar ways, and can be paraphrased in the same way. Note that the probabilistic account is a global account towards conditionals. It follows that conditionals are accepted or not with respect to a whole belief state. But that would mean that we can never learn anything from accepting a conditional. To account for the latter, it seems we have to assume that a conditional expresses a proposition. Most important, however, is that if we agreed with Gibbard, it would be much harder to explain, following the pragmatic tradition, the (seemingly) objective concept of counterfactuality in terms of the epistemic notion of conditional belief.

It seems that a unified account would mean to follow Stalnaker and use conditional logic for both subjunctive and indicative conditionals. But the threat of Gibbard's problem remains.

If we want to analyse all conditional sentences as propositions, have a uniform Ramsey test analysis of the conditional, and demand that revision should satisfy preservativity, we have to give up the assumption that the conditional has a fixed interpretation. The most straightforward and clarifying way in which this can be done is to follows Harper (1976).

5.7 Harper's principle and iterated revision

To account for iterated revision by learning new information, we would like our change function to obey all eight Gärdenfors postulates, and in particular ($K^*7,8$) should be satisfied. One of the nice things about the original Lewis/Stalnaker account is that nested conditionals, or iterated revision, do not give rise to interpretation problems. A similarity relation is given once and for all. A conditional, like any other sentence simply denotes a proposition. However, we have seen that revision by imaging does not guarantee that ($K^*7,8$) or its probabilistic variant will be obeyed. Harper (1975) tried to construct non-trivial models of iterated belief change by restricting Stalnaker's hypothesis to the level of certainty: $P(A > B) = 1$ iff $P(B/A) = 1$. Stalnaker (1976c) showed, however, that by making the assumption that ">" has a fixed interpretation, and by accepting the following limited version of (CSH): if $P(A \wedge C) \neq 0 \Rightarrow P(A > B)/C) = 1$ iff $P(B/A \wedge C)) = 1$, that is, by accepting ($K^*7,8$), the conditional connective collapses into material implication.[326]

As we have concluded earlier, if we want to analyse conditionals via the Ramsey test paradigm, we have to make the interpretation of the conditional dependent on the particular acceptance states. In § 5.6.6 we have already seen how that can be done: make sure that the selection function obeys the preservativity condition. However, to account for iterated revision, or nested indicative conditionals, this won't quite do. Even by accepting the preservativity condition it is still not guaranteed that the following more general constraint is met

($K^*7,8$) $C'_K(A) \cap B \neq \varnothing$ only if $C'_K(A \wedge B) = C'_K(A) \cap B$.[327]

But it is this constraint that is needed to handle iterative revision, and thus indicative conditionals nested to the right.

Van Fraassen (1976) was able to make the meaning of the conditional context dependent, and in principle still could account for embedded conditionals. However, he had to give up ($K^*7,8$) too. Thus, the question arises whether it is possible to account for iterated revision, and thus for embedded conditionals, without giving up ($K^*7,8$)?

Harper (1976) proved that we can, without the consequence that ">" is material implication. The price he had to pay, however, was that conditionals are even more context dependent than in Van Fraassen's construction. In Van Fraassen's theory, both of the connectives in a conditional like $A > (B > C)$ have the same meaning, while for Harper the two connectives have a different meaning. But exactly this made it possible to obey ($K^*7,8$). The way he built this context dependence into the meaning of the connective, into the selection function, is to make the Lewis/Stalnaker notion of similarity dependent on the information state. This dependence is made so systematic that iterated revision is not problematic anymore. Harper makes the meaning of the conditional context dependent by accepting the following principle, that I will refer to as Harper's principle (HP):

(HP) Only propositions decided by K should count in determining comparative similarity relative to K.[328]

---

[326] See also Gibbard (1980) for a closely related result.

[327] Consider the three logically independent propositions P, Q and R, and the eight worlds representing their possible combinations. Then we consider the following three propositions, $K = Q \cap R$, $A = \neg Q \cap (\neg R \cup P)$ and $B = \neg Q \cap (\neg R \cup \neg P)$. It can now be checked that $C'_K(A) = \neg Q \cap (P \equiv R)$, $C'_K(A) \cap B = \neg Q \cap P \cap R$, but $C'_K(A \wedge B) = \neg Q \cap \neg R$.

[328] Note that by accepting Harper's principle, Harper's account will be very close to the premise set accounts of Veltman (1976) and Kratzer (1981) developed around the same time, if no additional principle of measuring similarity (like *lumping*) is assumed.

Harper defends this principle as follows:

> If one reflects on the role of Ramsey test conditionals the new principle is very plausible. As an acceptance context the total content of K is given by the propositions it decides, therefore it is just these propositions that should form the basis of judgement of comparative similarity relative to K. (Harper, 1976, p. 130)

To formalise the principle, first a definition. For subsets S of W, belief states K, and worlds x and u, let $S^x{}_uK$ be the set of K-decided propositions in S on which x and u differ:

$$S^x{}_uK = \{A \in S: (K \subseteq A \text{ or } K \subseteq {\sim}A) \text{ and } ((x \in A \,\&\, u \notin A) \text{ or } (x \notin A \,\&\, u \in A))\}$$

Harper formalises Harper's principle in the following way (where $S \subseteq \wp(W)$):

(HP) If $S^x{}_uK = S^y{}_uK$ and $S^x{}_vK = S^y{}_vK$, then $u \leq_K{}^x v$ only if $u \leq_K{}^y v$

If u and v both differ from x on exactly the same K-decided propositions in S on which they differ from y, then their comparative similarity to x relative to K must agree with their comparative similarity to y relative to K.

Now we define a relational measure of nearness based on the assumption that only the propositions that are decided by K determine similarity. How is that done? We can say that u is at least as similar to w as v, iff the cardinality of the K-decided designated propositions on which u differs from w is less than or equal to the cardinality of the K-decided designated propositions on which v differs from w. In the simplest way we can take this set of designated propositions to be the set of atomic propositions, but you might also take this set to be any other set of arbitrary subsets of W.[329] More formally, we can define relative nearness in the following way: $u \leq_K{}^w v$ iff $|S^w{}_uK| \leq |S^w{}_vK|$, for any $w \in K$, where S is the set of designated propositions that potentially determine similarity. From this definition it follows that the similarity relation obeys weak centering,[330] transitivity, connectedness, the limit assumption and that Harper's principle is true. This similarity relation gives rise to a selection function and a system of spheres. The selection function is defined as follows: $f^K{}_w(A) = \{v \in A | \forall u \in A: u \leq_K v \Rightarrow |S^{w'}{}_uK| = |S^{w'}{}_vK|\}$, for any $w' \in K$. This selection function does not satisfy strong, but weak centering, $(w \in K \text{ and } w \in A) \Rightarrow w \in f^K{}_w(A)$. We know already that if the similarity relation obeys transitivity, connectedness and weak centering, the system of sphere model will look similar to Grove's model which he used to analyse belief revision. Now we can account for iterated revision, because if a set S of propositions that potentially determine similarity is assumed, from any set of possible worlds we can determine a system of spheres that belongs to it. Thus, we might say, qualitative revision is no longer a function from a system of spheres to a set of possible worlds, as in Grove (1986), but a function from a system of spheres to a system of spheres.[331]

Let the change function determined by accepting the Harper's principle be denoted by C". Then the following can be proved:

---

[329] See Harper (1976) for details. For instance, we can order the elements of S, first we look only at elements of S that correspond with lawlike sentences, and if that does not discriminate enough, we can also look at other propositions.

[330] Because only the K-decided propositions count in determining similarity, for any world w in K, $f^K{}_w(T) = K$. Strong centering cannot be assumed anymore. Thus, we can no longer infer $A > C$ from $A \wedge C$. According to Adams (1976) we shouldn't, with it we cannot account for *explanatory uses* of counterfactuals.

[331] For a more abstract account of iterated revision, see Spohn (1987).

(Theorem)
If $\leq_K{}^w$ is a relativised comparative similarity relation satisfying Harper's principle, and assume all the definitions given above, then

(a) $f^K{}_x(A) = f^K{}_y(A) = C''_K(A)$, if $x,y \in K$
(b) $f^K{}_w(A) \subseteq K$ if $K \cap A \neq \emptyset$ and $w \in K$ $(K^*3,4)$[332]
(c) $C''_K(A) \subseteq A$
(d) $C''_K(T) = K$ (where T is a tautology)
(e) $C''_K(A) \cap B \neq \emptyset$ only if $C''_K(A \wedge B) = C''_K(A) \cap B$[333]

And this gives us exactly what we wanted for indicative conditionals. C" is preservative, and the assumption that the similarity function is context independent is given up. In this way indicative conditionals are made heavily context dependent, without giving up the assumption that they express propositions and should be handled by the Ramsey test analysis. Let K be $\cap\Omega(P)$ for a particular Popper function P, very simplistically we can then define for all non empty K and A: $P(B/A) = |B \cap C''_K(A)| / |C''_K(A)|$. Suppose now simplistically that $P(A) = |C''_K(A)|/|C''_K(T)|$, then a relativised and weaker version of Stalnaker's hypothesis: $P(A >_K B) = 1$ iff $P(B/A) = 1$, is true for all Popper functions P.[334]

Note also that the original Lewis/Stalnaker notion of similarity is a special case of Harper's construction. For Lewis and Stalnaker, the set K that represents the belief state, is simply a singleton set. Thus, given a set S, $u \leq_w v$ iff $|S^w{}_u\{w\}| \leq |S^w{}_v\{w\}|$. Because a world decides all propositions, all propositions in S actually determine similarity. Lewis and Stalnaker always argued that the notion of similarity is context dependent. In our terms we might say that Stalnaker (1968) and Lewis (1973) already made the selection function dependent on what propositions potentially determine similarity, and that Harper showed that this selection function can also systematically depend on what is believed. Where Lewis and Stalnaker could already account for the fact that two agents whose beliefs are compatible with each other could justifiably assert two incompatible conditionals, because they assumed different ways of selecting closest worlds, Harper can also explain such cases by pointing to the difference of information available to the two agents.

5.8 Gibbard's problem revisited

Let's now go back to Gibbard's poker game example. To account for the conclusion in the poker game case that Pete folded is not so difficult. Let A, B ,C, D, E and F be the following propositions:

A: Pete called,
B: Pete won,
C: Stone's hand is quite good,
D: Pete knows Stone's hand as well as his own,

---

[332] Note that when $K = \{w\}$, strong centering follows.

[333] Note that if we make the assumption that belief states should be represented by sets of possible worlds and that $f^K{}_w(A) = \emptyset$ iff $A = \emptyset$ for all non-empty K and $w \in K$, it follows immediately that minimal revision governed by principle (2) satisfies the Gärdenfors postulates.

[334] Harper's result is not dependent on the particular way we defined probability. That the stronger result, $P((A >_K B)/T) = P(B/A)$, cannot be proved if the logic of A > B is Stalnaker's logic C2, is proved by Stalnaker (1976b), where he proves that Lewis's triviality result for Stalnaker's logic does not depend on the assumption that conditionals have a context independent fixed meaning.

E: Pete is disposed to fold on knowing that he had the losing hand, and
F: Pete had the losing hand.

Let I be our belief state. We know that Zack believes C, D and E, so $\forall w \in I$: $K(z,w) \subseteq C \wedge D \wedge E$. It follows that for all w in I, $C_{K(z,w)}(A) \subseteq B$. We also know that Jack believes C and F, so $\forall w \in I$: $K(z,w) \subseteq C \wedge F$. It follows that for all *w* in I, $C_{K(j,w)}(A) \subseteq \sim B$.

But how can we conclude from both assertions "If Pete called, he lost" and "If Pete called, he won", that Pete folded without knowing who made what statement? We assume that both are justified in claiming what they did, because the premises on which they base their conclusion are true and they believe that what they say is true. I know D, E and that Jack knows both hands. I argue as follows: Suppose Pete had the losing hand, by D he knows he has the losing hand, and by E he folded. Now suppose Pete had the winning hand, by D he knows he has the winning hand, so the conditional "A > B" would be true. Either Jack or Zack gave me a note which said "A > ~B". By looking at the context dependence of "A > ~B", we can not only determine what is expressed by the sentence once we know who wrote the letter, but once we know enough about the belief states of the possible writers and we assume some reasonable principles of communication, we might also be able to determine in what context we were, that is, determine who wrote the letter. Suppose the writer was Jack. I know that just like Pete, also Jack knows both hands. By the principle of non-contradiction this would be inconsistent with each other, so Pete could not have had the winning hand and knew it, so he folded. Now suppose it was Zack who gave me the note "A > ~B". I know that Zack knows that Pete knows both hands. I know also that Zack knows Stone's hand. Because Zack made his claim on the basis of Stone's hand, I conclude that Stone's hands are good. Because I know D and E, I conclude that Pete folded. Because I have considered all possible cases, and from all possible cases it followed that Pete folded, I conclude that Pete folded.

So, also without accepting the material implication account of indicative conditionals we can infer that Pete folded. But this was not the main threat of Gibbard's example. His example was meant to show that it makes no sense to claim that the respective conditionals express propositions, and that even if we don't know so much about the belief states of Jack and Zack we still can infer that Pete folded if we assume that the two messages are reliable. We have seen that Stalnaker (1975) made the selection function, and thus conditionals, context dependent. But that doesn't help as long as the meaning of the conditionals depends on the same context. What is needed is that the proposition expressed by the conditional depends on the beliefs of the *speaker*. Because the speakers can have different beliefs, the meaning of the same conditional sentence can still be different. But the problem is that sometimes we don't know who the speaker is, so, according to Gibbard, there can be no proposition expressed by an indicative conditional. But as Gibbard notes, the same thing can be true for sentences with indexicals. The proposition expressed by the sentence written on a postcard send without addressee with the message "I'm doing fine", depends in the same curious way on who the unknown sender of the postcard is.

This suggests that Gibbard's problem should be solved in the same way as an utterances which uses referential expressions, but for which it is not clear what the actual referent is, *diagonalisation*.

Let us first sketch the situation. In the initial situation for me, for Jack and for Zack, there are three possibilities, Pete folded, Pete called and won, and Pete called and lost. Let's call those three situations w_1, w_2 and w_3 respectively. As far as we know, if Pete calls he might either win or loose. After Zack and Jack looked into the cards, their information states changed. Assuming that the utterers of *If Pete called, he won* and *If Pete called, he lost* were justified in claiming what they did, their belief states can be represented by $\langle\{w_1, w_2\}, f\rangle$ and $\langle\{w_1, w_3\}, g\rangle$, respectively, where the two selection functions obey the

preservativity principle with respect to the belief state to which they belong, viz.: f_{w1}(Pete called) = $\{w_2\}$, and g_{w1}(Pete called) = $\{w_3\}$.

To account for diagonalisation we need to make a distinction between context and index. Contexts and indexes are different kinds of entities. A context consist of a certain aspect of a world. The context determines *what* is expressed by a sentence, while the index determines whether what is said is true or not. If someone writes me a note which says 'If A, then B', it depends on the context world what proposition is expressed by it. Two kinds of contexts are relevant in our example, one in which the belief state of the utterer of the conditional can be represented by $<\{w_1, w_2\}, f>$, and one in which the belief state of the conditional can be represented by $<\{w_1, w_3\}, g>$. Thus, different contexts correspond with different selection functions; *f* and *g*. I will assume that the same contexts can be 'part of' different worlds. If we don't know in what context we are, we don't know who slipped the note, but still want to determine what proposition is expressed by 'If A, then B', we diagonalise. We consider the set of context-world pairs in which the writer of the note in that context world wrote down a true proposition in the world. It is easy to see that *If Pete called, he won* is true with respect to the following context-index pairs: $<f,w_1>$, $<f,w_2>$ and $<g,w_2>$, while the assertion *If Pete called, he lost* is true with respect to the following context-index pairs: $<g,w_1>$, $<f,w_3>$ and $<g,w_3>$. What I learn when I accept both sentences is not who made what statement, nor whether the one or the other is true, I only learn that we have to be in world 1; only in w_1 both sentences can be true. I conclude that Pete folded.

There is a different, but related, way to account for Gibbard's problem without giving up the assumption that all conditionals state propositions. For all sentences whose interpretation depends on context, there are two ways in which two agents can disagree about its truth value. First, they can agree about what is said, but disagree about whether what was said is true, and second, they can have identical beliefs about the world in all relevant ways, but disagree about the truth of the sentence because they disagree about what is said. In case of conditionals the latter kind of disagreement can be accounted for by saying that the way to select nearest worlds differs. Even if I have all the relevant information of both Jack and Zack about the cards and the dispositions of Pete, there is both a way to think of *If Pete called, he lost* as being true and as being false. The conditional is true, if what determines similarity is what cards Pete and Mr. Stone have, if not the cards, but the dispositions of Pete determines similarity, the conditional is false. Let us say that selection function *f* goes with similarity by cards, and selection function *g* with similarity by Pete's disposition. Then the two propositions asserted by Zack and Jack are respectively $\{w \in W | f_w(A) \subseteq B\}$ and $\{w \in W | g_w(A) \subseteq -B\}$. Gibbard's problem is no threat to the principle of conditional non-contradiction as long as the latter is restricted to the proposition expressed in a fixed but arbitrary context, because Zack's and Jack's use of respectively the sentences *If Pete called, he won* and *If Pete called, he lost* simply do not express contradictory propositions. We saw already how to account for the fact that from their respective claims I can conclude that Pete folded.

The two ways to account for Gibbard's problem correspond with the two ways in which the meaning of '>' depends on context. According to the diagonalisation solution the two statements are not contradictory because the belief states of the two agents are different. According to the second solution the reason is that the propositions that potentially determine similarity are different from each other. But both proposals have the following in common: What is expressed by a counterfactual sentence is functionally dependent on the intention of the speaker; the criteria for selecting nearest possible worlds. If we say that the intention of the speaker is the relevant contextual factor, we can say that the character expressed by *if A, then B* is $\lambda f.\{w \in W | f_w(A) \subseteq B\}$.

5.10 Subjunctive conditionals again

According to the Ramsey test analysis, A > B is accepted in K iff B is accepted in K revised by A. The triviality results showed why this analysis is not as obviously true as it was hoped at one time. However, the problem posed by the triviality results can, at least formally, be accounted for by making the conditional context dependent. We have seen that this was proposed by Harper and Morreau. But Harper (1976) did not claim that his analysis of conditionals should be used for all kinds of conditionals, in particular, that it should be used for the analysis of counterfactuals. He argued that the analysis should only be used for those conditionals that more or less reflect the conditional beliefs of the agents who utter them. There are various reasons to think why subjunctive conditionals should not be handled in this way. First, as observed by Adams (1975), an analysis of counterfactuals in terms of the actual conditional beliefs of the agent cannot account for certain *explanatory uses* of counterfactuals. As noted by Stalnaker (1975), there are subjunctive conditionals whose antecedents are consistent with what is presupposed, but for whose interpretation we necessarily should look outside the context that represents this common background knowledge. He suggests that this is exactly the reason why we use the subjunctive mood. In a sentence like *If Mary were allergic to penicillin, she would have exactly the symptoms she is showing* the conditional is presented as evidence for the truth of its antecedent. If subjunctive conditionals are handled via the epistemic Ramsey test analysis, and if the relevant context is that what is currently presupposed, the sentence would be trivially true and so could be no evidence for the truth of the antecedent.

The most convincing reason why subjunctive conditionals should not be analysed via the most straightforward reading of the Ramsey test analysis is of course that it becomes unclear how we could account for the difference between Adams' Oswald-Kennedy examples when they are stated in indicative and subjunctive mood. In a similar way it becomes impossible to account for the following:

"Suppose I accept that if Hitler had decided to invade England in 1940, Germany would have won the war. Then suppose I discover, to my surprise, that Hitler did in fact decide to invade England in 1940 (although he never carried out his plan). Am I now disposed to accept that Germany won the war? No, instead I will give up my belief in the conditional. In this case, my rejection of the antecedent was an essential presupposition of my acceptance of the counterfactual, and so gives me reason to give up the counterfactual rather than to accept its consequent, when I learn that the antecedent is true". (Stalnaker, 1984, p. 105).

Let A and B be respectively *Hitler decided to invade England in 1940*, and *Germany would have won the war*. As Gärdenfors (1988) noted, to account for this example in the global epistemic approach towards revision, that is, giving up A > B rather than to accept B as a response of learning A, it is needed that ~B is stronger entrenched that A > B. The problem is that this cannot be the case. According to the epistemic account, conditionals do not really express propositions. They are only accepted or not in a whole belief state represented by something like a system of spheres. My conclusion is that at least some counterfactuals must denote a proposition. But that some counterfactuals must denote a proposition doesn't mean that they are thus context independent. We have seen already that we can say that even indicative conditionals express propositions, although what is expressed by such indicative conditional sentences is very context dependent. what is expressed by indicative conditionals is extremely context dependent because it not only depends on the speakers criteria for selecting, but also the particular belief state of the speaker. The proposition expressed by subjunctive conditional sentences is not so extremely context dependent, it seems that only the propositions that potentially determine nearness depends on context. But as we have seen in §5.8, this is already enough to make it possible that even two subjunctive conditional sentences of the form *If A were the case, B would be the case* and *If A were the case, ~B would be the case* can both be true at the same time.[335]

[335] See chapter 7 of Stalnaker (1984) for more discussion.

From now on we will say that counterfactuals are simply true or false in a world according to a contextually given selection function. If this is so, counterfactuals express propositions and can thus be less strongly entrenched than their consequent. In particular for the Hitler example, it becomes possible now that the counterfactual A > B is given up because learning A, that Hitler decided to invade England in 1940, does not result in giving up my belief in ~B, that Germany lost the war.

It seems we have come to the same conclusion as Lewis (1973, 1975a) and Gibbard (1980): the Ramsey test is relevant for the analysis of indicative conditionals, but this is not the case for counterfactuals. But their position leaves an important question to be answered: if the selection function is not to be explained as the projection of a methodological policy onto the world, how then should we understand the meaning and role of counterfactuals?

We have seen convincing arguments why belief in counterfactuals should not be explained in terms of conditional beliefs in the most straightforward reading of the Ramsey test analysis. But that does not mean that the project of explaining the meaning of counterfactuals in terms of conditional beliefs is completely hopeless.

> If we could distinguish and filter out those aspects of our epistemic situation which derive more from our parochial perspective and less from the way we take the world to be, we might be able to explain the acceptance of conditional propositions in terms of the open conditional that would be acceptable in idealised contexts which abstract away from those aspects. (1984, pp. 115-116)

From those suggestions to an account of the connection between beliefs in counterfactuals and conditional beliefs is a long way, and I have nothing to offer. It is clear what should be accounted for, the fact that we normally understand each other if we use counterfactuals. This doesn't mean that the counterfactual connective thus has a fixed interpretation, not even if the set of propositions that potentially determine similarity is fixed. It still depends on what is accepted. But this acceptance state need not be the actual belief state or presupposition state of the agent. That the use of counterfactuals does normally not lead to interpretation problems suggests that in most uses of counterfactuals it is relatively clear what criteria for selecting is assumed by the utterer. This, in turn, means that the *pragmatics* of conditionals must be quite systematic. It is possible that for some uses of subjunctive conditionals the selection function reflects the current belief state of the utterer, sometimes his prior belief state,[336] sometimes an information state that is simply consistent with natural laws, and sometimes something else. Thus, I believe that the meaning of the conditional connective, '>', should in the end be explained in terms of conditional beliefs, but the relevant belief state can be a prior belief state, or an information state that reflects the beliefs of a great number of agents, or maybe a combination of both. Of course, once it is assumed that the meaning of the connective '>' should be explained in terms of conditional beliefs, and if conditional beliefs are interpreted by conditionalisation or qualitative variants thereof, the results of Stalnaker (1970a) and Van Fraassen (1976) might be relevant again.

The pragmatics of conditionals starts with the assumption that the selection function is context dependent. In general it is difficult for counterfactuals to say more about in what way the selection function is determined by context. One pragmatic aspect about conditionals, however, is pretty clear. This is the way the selection function changes during an argument. To that we will turn now.

---

[336] That we should look at a prior state in one way or another (i.e. a prior belief state, or a prior state of the world) for the analysis of counterfactuals has been proposed by a number of people. Adams (1976) was probably the first, followed by for instance Thomason & Gupta (1980), Lewis (1979c) and Skyrms (1980a/b). Thomason & Gupta suggest that looking at current versus a prior state is all there is to the distinction between Adams' Oswald-Kennedy examples in respectively indicative and subjunctive form. Lewis (1979c) argued that the notion of similarity is not as vague as has been suggested by Fine (1975) (and later by Kratzer, 1989), if it is recognised that for determining similarity, prior states of the world are crucial. I don't believe though that looking at a prior state can be everything there is to the subjunctive mood.

5.10 Invalidity explained by illegitimate change of context

The meaning of a conditional depends on the way similarity is measured. If a speaker asserts a (subjunctive) conditional, he has a certain way of selecting similarity in mind. If the selection function used for the interpretation of counterfactuals can be dependent on the speaker's intention, a hearer can disagree in two ways with the speaker with respect to the truth-value of a counterfactual. It can be the case that the hearer understood the speaker correctly and that they disagree about the facts. But, as in other cases of context dependence, it is also possible that the hearer has misunderstood the speaker's intention. He disagreed with the speaker because he assumed a different way of selecting nearest possible worlds, he picked out the wrong selection function. He misunderstood the speaker because he assumed that a different proposition was expressed than what the speaker actually wanted to express.

What is problematic about the analysis of conditionals is not only that it is difficult to determine what the relevant set of propositions is that determines similarity, but also that this set should stay stable during an argument involving more conditionals. In inferences where in the middle of the argument the set of propositions that determine similarity is changed, a fallacy will arise. Let S and S' be two sets of propositions that potentially determine similarity. Let us say that if first S and then S' measures similarity, a *context change* has occurred. In principle the set S' can stand in four kinds of relations to S; S' can be independent of S, S' can be a subset or a superset of S, and finally S and S' can be disjoint.

We have already seen some examples (the examples from Gibbard and Tichy) where S and S' do not stand in a sub- or superset relation to each other. In those cases, with the same conditional sentence something completely different can be meant. In more interesting cases, S and S' *do* stand in an inclusion relation in one way or another, and the selection function that corresponds to one set is thus more fine grained than the selection function that corresponds with the other.

Let us first look at a case where the set of propositions that determine similarity decreases during the argument. In these cases a context change occurs, but we typically find it difficult to detect this change of context. Some famous fallacies typically arise in these kind of circumstances. Consider the following argument for fatalism of Dummett:

> Either I will be killed in this raid or I will not be killed. Suppose that I will. Then even if I take precautions I will be killed, so any precautions I take will be ineffective. But suppose I am not going to be killed. Then I won't be killed even if I neglect all precautions; so, on this assumption, no precautions I take will be either ineffective or unnecessary, and so pointless. (from Stalnaker, 1975)

The argument is of the following form: $K \vee \sim K$, (if K, then (if P then K) thus Q), (if ~K, then (if ~P, then ~K), thus R), thus Q or R. The argument is invalid, because the statements 'If P, then K' and 'If ~P, then ~K' are not valid. But, as Stalnaker points out, in the contexts in which these conditionals are used (respectively K and ~K), they give rise to reasonable inferences (for the notion of reasonable inference, see above). The problem with the argument, according to Stalnaker, is that it assumes that the conclusion is a reasonable inference given that the sub-arguments are reasonable inferences. But that is not the case. This is only the case if all the sub arguments are reasonable inferences with respect to the *same* context, which was not the case in the fatalism argument. The conditionals used in the sub arguments are true in the contexts where respectively K and ~K are accepted. But the conditionals can no longer be accepted in the main context, a context where neither K nor ~K is accepted.

To illustrate, consider the following belief state: <W,f>, where W = $\{w_1, w_2, w_3, w_4\}$, K = $\{w_1, w_2\}$, P = $\{w_2, w_3\}$, Q = P > K, R = ~P > ~K, $f_{W1}(P) = f_{W2}(P) = f_{W3}(P) = f_{W4}(P) = P$, $f_{W1}(\sim P) = f_{W2}(\sim P) = f_{W3}(\sim P) = f_{W4}(\sim P) = \sim P$.

It is clear that in this belief state, Q ∨ R is not true, so the inference is not valid. But because in <W,f> the preservativity condition is satisfied, the inference is not even reasonable. Still, the sub arguments are reasonable because if you assume K and preservativity, you end up in belief state <K, g>, where $g_{W1}(P) = g_{W2}(P) = \{w_2\}$, and if you assume ~K and preservativity, you end up in belief state <~K, h>, where $h_{W3}(\sim P) = h_{W4}(\sim P) = \{w_4\}$.

We have seen that in analysing conditionals in discourse or argument, we easily go from a more to a less determined selection function. The less determined the selection function is, the more worlds in which the antecedent is true we have to check as to whether the consequent is true. Thus, the more difficult it will be for a conditional to be true.

5.11 The systematicy of context change

Lewis and Stalnaker recognised the context dependence of the selection function for the analysis of counterfactuals. What they did not so clearly see, I think, is that this context dependence is in some cases very systematic. An important argument for both was that counterfactuals of the form A > C and A ∧ B > ~C can be true simultaneously. However, as noted by Annette Frank (1997), only discourses of the form "A > C, and A ∧ B > ~C" are acceptable, the same discourse in the converse order is out. But if only truth mattered, Lewis and Stalnaker would not predict there being a difference. We can conclude that the acceptability of a counterfactual does not only depend on whether or not it can be true in a given context. It seems that the interpretation of a counterfactual changes the context in such a way that other kinds of counterfactuals can no longer be appropriately uttered in the new context. Can we build this context change into our theory such that counterfactuals are still analysed via selection functions? I think we can, but we need some more apparatus.

First I will introduce a function, k, closely related to Spohn's (1987) ordinal functions. Let S be a set of propositions that (potentially) determine similarity. As always, once we have a set S, we can determine a similarity relation and a selection function:

$u \leq_w v$ iff $|S^u{}_w\{w\}| \leq |S^v{}_w\{w\}|$

$f_w(A) = \{v \in A \mid \forall u \in A: u \leq_w v \Rightarrow u = v\}$

Now we let $k^S{}_u(w)$ stand for the number of propositions in S on which w and u differ in truth value, thus $k^S{}_u(w) =_{df} |S^w{}_u\{w\}|$. In terms of this, $k^S{}_u(A)$ is defined as $\min\{k^S{}_u(w) \mid w \in A\}$. To account for the way a counterfactual changes a context, I limit myself to consistent context changes. Thus if a sentence A is interpreted to be false in all possibilities of the context I, the context change of I by A, [[A]] (I), will simply be the empty set. Normally in propositional logic a context is represented by a set of possible worlds, and [[A]] (I) is I ∩ A. This simplification will be given up. I will represent a context by a set of pairs like <w,S>, where S is the set of propositions that determines similarity used for the analysis of counterfactuals. If we define I^* and I(w) as follows,

$I^* = \{w \in W \mid \exists S: \langle w,S\rangle \in I\}$

$I(w) = \{S \subseteq \wp(W) \mid \langle w,S\rangle \in I\}$,

I demand that for each A, [[A]] is an element of $\{<I,I'>| \forall w,w' \in I^*: I(w) = I(w') \ \& \ \forall w,w' \in I'^*: I'(w) = I'(w')\}$. If A is not a conditional, and there is no conditional that occurs in A, [[A]] (I) is simply $\{<w,S> \in I: w$ makes A true$\}$. However, if A is of the form B > C, the context changes in a less simplistic way:

$$[[A > C]](I) = \{<w,S>| \exists S' \supseteq S: <w,S'> \in I \ \& \ \forall B \supseteq A: k^S{}_u(B) = k^S{}_u(A) \ \& \ \forall S'' \supseteq S[\forall B \supseteq A: k^{S''}{}_u(B) = k^{S''}{}_u(A) \rightarrow S'' \subseteq S]\}$$

What intuitively is going on is that a counterfactual influences the fine grainedness of the selection functions with respect to which other counterfactuals are interpreted. In particular, if I is a context in which the counterfactual A > C is accepted, and if $A \subseteq B$, then for each $<w,S> \in I$, the set of closest B-worlds to w, $\{v \in B| \forall u \in B: |S^u{}_w\{w\}| \leq |S^v{}_w\{w\}| \Rightarrow u = v\}$, will contain at least also some A-worlds. This has the effect that it is predicted that the discourse "A > C and $A \wedge B$ > ~C" is fine, but "$A \wedge B$ > ~C and A > C" is not, just as we desired.

One thing is problematic about the above solution, though. By the way we have built the context change induced by counterfactuals into the semantics, we have to give up the uniqueness assumption. If we don't want to use supervaluation, the above suggestion spoils the analysis of *only if* clauses. Maybe this shows that we did things in the wrong way, context doesn't make the selection function less determinate, but only allows for more possible selection functions.[337] However, there is another way to account for the context change induced by counterfactuals that can account for *only if* clauses. Let me first give an intuitive motivation for this account.

It seems reasonable that any adequate theory of counterfactuals must account for the fact that at least most of the time instantiations of the following formula (Simplification of Disjunctive Antecedents, SDA) are true:

(SDA) $[(A \vee B) > C] \rightarrow [(A > C) \wedge (B > C)]$

The problem is that if we make this principle valid, by saying that $f_w(A \vee B) = f_w(A) \cup f_w(B)$, the theory looses one of its most central features, its non-monotonicity. The principle of monotonicity,

(MON) $[A > C] \rightarrow [((A \wedge B) > C)]$,

becomes valid. That is, by accepting SDA, we can derive MON on the assumption that the connectives are interpreted in a Boolean way.[338] The Lewis/Stalnaker account does not validate MON because SDA is not a theorem of their logic. The same is true for Adam's probabilistic account. However, are those who claim that SDA should be a theorem not right? It certainly is the case that from (10) we infer (11) and (12):

(10) If Spain had fought on either the Allied side or the Nazi side, it would have made Spain bankrupt.
(11) If Spain had fought on the Allied side, it would have made Spain bankrupt.

---

[337] The reason is that each possibility, <w,S>, of a context, I, corresponds with a possibility <w,f>, where f is a Lewis selection function, and with a *set* of possibilities $\{<v,f_S>| v = w \ \& \ \forall A \subseteq W: f_{S,w}(A) \subseteq f_w(A)\}$, where each f_S is a selection function that satisfies the uniqueness condition.

[338] From A > C and the assumption that connectives are interpreted in a Boolean way, we can derive $((A \wedge B) \vee (A \wedge \sim B)) > C$. By SDA we can then derive $(A \wedge B) > C$.

(12) If Spain had fought on the Nazi side, it would have made Spain bankrupt.

Contrary to Lewis and Stalnaker, the inferences are predicted to be valid by a strict conditional account.[339]

5.12 A variable strict conditional account

The oldest way to account for the peculiarities of counterfactual conditionals, was to interpret them as modalised material conditionals. Thus, if "$\rightarrow$" denotes the material conditional, $A > B$ is true in w iff $A \rightarrow B$ is true all worlds w' that are accessible from w. The strict conditional account predicts that transitivity, strengthening of antecedent, and contraposition are all valid.[340] Stalnaker (1968) and especially Lewis (1973) argued that counterfactuals cannot be analysed as strict conditionals, because in that way we cannot account for certain fallacies. In particular, counterfactuals do not behave in a monotone way and don't obey transitivity and contraposition. And they are right, if the accessibility relation stays constant, a strict conditional account will not do. According to Lewis and Stalnaker (and Adams), we should give a semantic account for the fallacies associated with counterfactuals. However, we have seen that to account for other fallacies, both take the relevant selection function to be context dependent, and that this context dependence seems to change systematically in a discourse. But if the relevant selection function sometimes systematically has to change during an argument, does the Lewis/Stalnaker account still has an advantage above a strict conditional account if we allow the accessibility relation to change during an argument? That all depends on how the accessibility relation is defined, how it can change during an argument, and how straightforward a strict conditional analysis can account for the fallacies associated with counterfactuals.

In a very interesting article, Warmbrod (1981) argued for something like a strict implication analysis, and to account for all kinds of fallacies related to counterfactuals by the principle that the relevant accessibility relation is not allowed to change during the argument. Remember that according to the strict implication account a counterfactual *if A then B* denotes the following proposition: $\{w \in W: \forall w' \in W[wRw' \Rightarrow w' \in (A \rightarrow B)]\}$, where R is an accessibility relation and $\rightarrow$ is material implication, i.e.: $(A \rightarrow B) = (\sim A \cup B)$. Equivalently, this is just $\{w \in W: R(w) \subseteq (A \rightarrow B)\}$, the selection function is replaced by an accessibility relation. But Warmbrod stays very close to the Lewis/Stalnaker account by basing the accessibility relation on a context dependent notion of similarity. In the following analysis I will stay close to the spirit of Warmbrod's analysis, although it is not exactly the way he himself implements his ideas. First, I assume just as before that there is a set of propositions S that determines a similarity relation, and that for any proposition A, $k^{S}{}_{w}(A)$ is defined as above. Let T be a set of propositions. Let's define max(T,S, w) and the on T dependent accessibility relation $R^{S}{}_{T}(w)$ as follows:

$$\max(T,w,S) = \{A \in T \mid \forall B \in T: k^{S}{}_{w}(B) \leq k^{S}{}_{w}(A)\}$$

---

[339] With Fine (1975), I don't think that this means that counterfactuals with disjunctive antecedents falsify the Lewis/Stalnaker account. The reason is that we cannot conclude *If A, then C* from all instantiations of conditionals of the form *If A or B, then C*: "If Spain had fought on either the Allied side or the Nazi side, it would have fought on the Nazi side. Thus, if Spain had fought on the Allied side, it would have fought on the Nazi side." (McKay & Van Ingwagen, (1977)).

[340] The three principles are closely related to each other (see Stalnaker, 1984): From *transitivity* to *strengthening of antecedent*: Immediate, if $A \wedge B > A$ is assumed to be valid. From *contraposition* to *strengthening of antecedent*: Assume weakening the consequent (if C is entailed by B, then $A > C$ is entailed by $A > C$). Suppose $A > C$, by contraposition $\sim C > \sim A$, by weakening the consequent $\sim C > \sim(A \wedge B)$, by contraposition $A \wedge B > C$.

$R^S_T(w) = \{u \in W \mid \forall v \in \cap max(T,w,S): u \leq_w v\}$

Thus, max(T,w,S) stands for the set of propositions in T that are least similar to *w* if the propositions in S determine similarity, and $R^S_T(w)$ stands for the set of worlds closer or equal to *w* than the worlds in the intersection of max(T,w,S). We will make some kind of limit assumption for $R^S_T(w)$: for all $A \in T$: $R^S_T(w) \cap A \neq \emptyset$. We will say that $A > C$ is true in *w* with respect to S and T iff $A \rightarrow C$ is true in all worlds in $R^S_T(w)$. From now on we leave the set S implicit. Two extra constraints should be obeyed, too. If we interpret several conditionals that intuitively 'belong together', we should analyse all those conditionals with the help of the same accessibility relation. If we analyse an argument to which a set of conditionals belong, we have to analyse those conditionals with respect to the same context, the accessibility relation is not allowed to change during the argument. The second extra requirement is that all the antecedents of the conditionals should be consistent with the set of accessible worlds. So, if we analyse a set of conditionals in a possible world w, we assume a contextually determined set of worlds, $R^S_T(w)$ with which all the antecedents of the conditionals are consistent. The easiest way to do that is that given our limit assumption, we assume that all antecedents of this set of conditionals are elements of T. All cases that have been assumed to be counterexamples to transitivity, contraposition, monotonicity, and SDA are special in that there is no single accessibility relation such that all the corresponding material implications of the premises are true in all the accessible worlds to *w* in which the extra constraints are satisfied. It follows that a strict conditional account is not so bad as is suggested by Lewis (1973). That the (apparent) counterexamples to monotonicity can be explained away as suggested above is clear,[341] but it is also true in case of the other three principles. Consider an (apparent) counterexample to transitivity:

(13) If Bush had not lost the election in 1992, Clinton would not have been President in 1994.
(14) If Bush had died during the Golf war in 1990, he would not have lost the election in 1992.
(15) If Bush had died during the Golf war in 1990, Clinton would not have been President in 1994.

According to Warmbrod, this would not be a counterexample to transitivity, because there is no (reasonable) set of worlds R(w), equally similar to the actual world *w* such that R(w) is consistent with the antecedents of (13), (14) and (15), and the material conditionals corresponding with (1) and (2) are also true in all worlds in R(w). The three sentences can only be true together, according to Warmbrod, if we change the context during the argument. But that is not allowed. In the first premise we assume that in all the worlds of the sphere suggested by the antecedent it is true that Bush was running for President in 1992, while this cannot be true in the worlds verifying the antecedent of (2). For contraposition we argue in the same way. Consider the following story suggesting that contraposition is not valid for counterfactuals:

> My dog is a mutt. His paternity is in some doubt, but even if his father were a purebred dog, my dog would still be mutt since his mother was one. Now consider the contrapositive of the conditional claim made in this remark: if my dog were a purebred, his father would be a mutt. (I assume that *mutt* and *purebred* are contradictory properties, as applied to dogs.) This conditional is not only false, but impossible, and so cannot be a consequence of the true conditional claim made in the story. (Stalnaker, 1984, p. 124)[342]

---

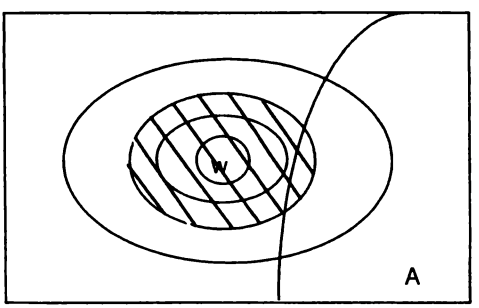
[341] When $A \rightarrow C$ is true in all accessible worlds, and intuitively also $(A \wedge B) > \neg C$ is true, there will be no accessible world in which $A \wedge B$ is true.

[342] Stalnaker (1984) assumes here, rightly I think, that *even if* conditionals are to be explained compositionally in terms of the meanings of *even* and of *if*.

The premise of the argument assumes it to be settled that the mother of my dog is a mutt. In that case, even if his father were not a mutt, my dog would still be a mutt. The antecedent of the conditional in the conclusion, however, states that my dog is not a mutt. But that can only be the case if neither his father nor his mother is a mutt. Because in the context we assumed for interpreting the premise it was assumed that the mother of my dog was a mutt, the invalid inference is not due to the invalidity of contraposition, but because of illegitimate context change.

We will see in a moment how a strict conditional analysis can account for apparent counterexamples to SDA, but first we will show how close Warmbrod's account can be to the LS account.

Let us assume that if we analyse a conditional of the form $A > C$ out of context, then $T = \{A\}$. In that case, the set of worlds accessible from world w, $R_{\{A\}}(w)$, will be the set of worlds that are at least as similar to *w* as those A-worlds that are most similar to w. If $R_{\{A\}}(w)$ is the relevant accessibility relation for antecedent A, then $w' \in R_{\{A\}}(w)$ iff $A \cap R_{\{A\}}(w) \neq \emptyset$ and $\forall w'' \in A$: $w' \geq_w w''$. Here is a picture:



From now on we abbreviate $R_{\{A\}}(w)$ as $R_A(w)$. Note that $R_A(w)$ satisfies weak, but not strong centering, and that factuality, $R_A(w) \subseteq A$, doesn't hold. The selection function can now be defined in such a way that it gives us the same worlds as Lewis and Stalnaker predict: $f_w(A) := R_A(w) \cap A$. Thus, if we fix a set of propositions that determine similarity and assume that in Warmbrod's account $A > C$ is true in w iff $R_A(w) \subseteq A \rightarrow C$, Warmbrod's strict conditional account is equivalent to the Lewis/Stalnaker account. Any motivation for the LS account also works for Warmbrod's account.

Now consider an (apparent) counterexample to SDA:

(16) If Spain had fought on either the Allied side or the Nazi side, it would have fought on the Nazi side.
(17) Thus, if Spain had fought on the Allied side, it would have fought on the Nazi side.

This counterexample too is easily explained by illegitimate context change. Let the premise have the form $(A \vee B) > B$, and the conclusion the form $A > B$. If w is a world in which (16) is true, $R_{A \vee B}(w)$ will contain some B worlds, but no A-worlds. However, this cannot be the same accessibility relation as the one needed to interpret (17). The claim that SDA is valid can be maintained if we are allowed to explain the counterexample away by an illegitimate change of the accessibility relation.

Consider the Bush-Clinton counterexample (13) - (15) to the validity of the hypothetical syllogism again. We have explained the counterexample away by saying that the accessibility relation could not stay constant during the argument. Note that the order of the apparent counterexample to transitivity is as follows: $B > C$ and $A > B$, thus $A > C$. It is remarkable that if we reversed the order of premises we wouldn't be so easily inclined anymore to accept the argument as a counterexample, because it would be more difficult to accept $B > C$ as true. A strong point about Warmbrod's pragmatic account is that he can explain why this is the case. If we analyse a set of conditionals we choose an accessibility relation such that all antecedents are consistent with the set of accessible worlds. If we analysed the conditionals $A > B$ and $B > C$ separately in *w*, the set of accessible worlds $R_B(w)$ will be smaller than the set $R_A(w)$, $R_B(w) \subseteq R_A(w)$, because B-worlds are not as far remote from actuality as A-worlds are. Thus if we first state $A > B$, the accessibility relation will be such that the antecedent of $B > C$ will be consistent with $R_A(w)$, we don't have to change the accessibility relation. However, in this case $B > C$ will probably not count as true. It is not at all clear that Clinton wouldn't have been president in 1994, if Bush wouldn't have been the republican candidate. If, as we actually did, first state $B > C$ and only then $A > B$, the counterexample seems convincing because we have no difficulty in considering both premises true. According to Warmbrod, the reason is the following. In interpreting first $B > C$ we consider the set of worlds $R_B(w)$. In this context it is accepted that Bush was the republican candidate for presidency, so that not loosing the election means winning it. But in all the worlds in this context the material implication corresponding to the next conditional would be trivially true. Because it is a conversational rule that all assertions should be informative, we have to change the context. We have to consider (also) worlds less similar to the actual world, where Bush died in 1990 and was not the republican candidate in 1992. It seems that we don't have big problems in going from a relatively small set of worlds to a bigger one to check whether a counterfactual is true. Note that how bigger the set of possible worlds is that we have to check, the more difficult it will be for the conditional to be true. This seems to be true in general, in one discourse we can easily strengthen the requirements or standards of precision, but we are not so easily inclined to weaken them. If Y is flatter than X, we can say 'X is flat, but Y is flatter', but saying 'Y is flat, but X is also flat' is much harder.[343]

There is another thing to be explained. By the appropriateness conditions of Warmbrod, we can explain why the discourse 'If this match were struck, it would light, but If this match had been soaked in water overnight and it were struck, it would not light' is no

---

[343] cf. Lewis (ms.) ' To say that X is known, I must be properly ignoring any unelimated possibilities in which not-X; whereas to say truly that Y is better known than X, I must be attending to some such possibilities. So I can't say both in a single context. If I say 'X is known, but Y is better known', the context changes in mid sentence: some previously ignored possibilities must stop being ignored. That can happen easily. Saying it the other way round --'Y is better known than X, but even X is known' -- is harder, because we must suddenly start to ignore previously *un*ignored possibilities. That can't be done, really'.

counterexample to monotonicity. In a discourse of the form 'If A, then C, but if A and B, then not C' the context changes during the argument. But it also has to be explained why the inverse discourse, 'If A and B, then not C, but if A, then C' is so much worse. First, note that in general $R_A(w) \subseteq R_{A \wedge B}(w)$, the requirement to be consistent with both A and B is bigger than the requirement of just being consistent with A. If we first state $A \wedge B > \neg C$ out of context, the relevant set of accessible worlds will be $R_{A \wedge B}(w)$. Because A is consistent with this set, for interpreting $A > C$ we don't have to change the accessibility relation. Instead, the conditional $A > C$ will by the strict conditional account simply be false. If we state the conditionals in the inverse order, however, the first accessibility relation will be $R_A(w)$. In all apparent counterexamples to strengthening of the antecedent, $A \wedge B$ will not be consistent with $R_A(w)$. We have to consider more possible cases and so make the conditional be more specific. With respect to this changed context it might very well be that the strict conditional $\Box(A \wedge B \rightarrow \neg C)$ will be true.[344]

In the beginning of this chapter we saw that *only if* constructions were an extra motivation for making the uniqueness assumption. However, when we have built the context change induced by counterfactuals into the semantics, we had to give up this assumption. Now I want to show that a strict conditional account can handle *only if* constructions. The argument is due to Zimmermann (personal communication). Just like Von Fintel (1994), we want to analyse *only if* constructions in terms of the meaning of *only* and *if*. The analysis of *if* is the strict conditional account. In distinction with the usual analysis of *only*, Zimmermann argues on independent grounds that *only A* denotes the following proposition: $\{w \in W \mid \forall B \supset \emptyset: w \in B \Rightarrow A \cap B \neq \emptyset\}$. Let us now assume, just as before, that normally in *B, only if A* it is A that is focused. In these cases, if R is the relevant accessibility relation, *B, only if A* denotes the following proposition: $\{w \in W \mid R(w) \cap A \subseteq B$ and $\forall D \supset \emptyset: R(w) \cap D \subseteq B \Rightarrow A \cap D \neq \emptyset\}$. Now we have to show that from *B, only if A* we can derive $B > A$. So we have to show that for any $v \in R(w) \cap B$, $v \in A$. Here is the proof: Let *v* be an element of $R(w) \cap B$. In that case, $\{v\} \cap R(w) \subseteq B$. But then it has to be the case that $A \cap \{v\} \neq \emptyset$, and thus $v \in A$.

---

[344] Nute (1984) has noted a problem for Warmbrod's account. If counterexamples to valid inferences according to the strict conditional account are to be explained away by illegitimate context change, it seems natural that we can always make up such counterexamples if we change the context in the middle of the argument. The following principle is valid according to the strict conditional account: $[(A > C) \wedge (B > C)] \Rightarrow [(A \vee B) > C]$. Suppose now that $B \cap R_A(w) = \emptyset$ and $R_A(w) \subseteq R_B(w)$. In that case we would change the context in the middle of the argument. But in this changed context it is not at all necessary that $R_B(w) \cap A \subseteq C$, and thus that $R_B(w) \cap (A \vee B) \subseteq C$. However, it seems hard to find counterexamples to the above principle.

Chapter 6

Some other attitudes

6.1 Introduction

In chapter 3 and 4 of this dissertation I tried to account for anaphoric and presuppositional dependencies across belief attributions. In the first part of this chapter I want to extend this analysis to attitudes of doubt and desire. To do that, we have to ask for each attitude how we should interpret it such that the *anaphoric* relations are handled well, and such that the interpretation rule accounts for the right *logical* relations. In the later parts of this chapter, the analysis is getting less formal and more suggestive. There I merely want to suggest how we might use the notion of entrenchment used in the last chapter, and the double indexing counterpart theory stated in the first chapter to analyse a number of other attitudes.

Before I start to discuss the attitude constructions of *doubt* and *desire*, let me first state the interpretation rule for belief sentences. In this chapter I won't be concerned with anaphoric relations in cases that more agents are involved, and will also ignore *de re* attributions of those attitudes, so I forget about counterpart functions. I will assume that belief states can be modelled by sets of possible worlds, and that belief attributions should be interpreted as in § 3.9, except that I will also ignore speaker's reference, and that indefinites can also introduce properties. As a result of this, belief attributions will be analysed as follows:

$$[[\mathrm{Bel}^{p}_{q}(a, A)]](S) = \{\langle g',h',w\rangle \mid \exists g,h,h'': \langle g,h,w\rangle \in S \;\&\; \forall w' \in K(\mathbf{a},w): \langle g',h'',w'\rangle \in [[A]](\{\langle k,l,w''\rangle \mid k = g \;\&\; l = h \;\&\; w'' \in g(p)\}) \;\&\; h''[q]h' \;\&\; h'(q) = W([[A]](\{\langle k,l,w''\rangle \mid k = g \;\&\; l = h \;\&\; w'' \in g(p)\}))\}$$

Now we are ready to discuss the attitude *doubt*.

6.2 Doubt

The interpretation of *doubt that* should be such that it is not closed under implication, but instead obeys *addition*, (A), *negative simplification*, (NS), and *downward entailment*, (DE):

A:	a doubts that A	$\Rightarrow$	a doubts that $(A \wedge B)$
NS:	a doubts that $(A \vee B)$	$\Rightarrow$	a doubts that A
DE:	a doubts that A & $B \subseteq A$	$\Rightarrow$	a doubts that B

Note that A and NS are both special cases of DE, so if we can give an interpretation rule that accounts for DE we seem to be ready.

Before we give the interpretation rule for *doubt* let us first look how it should behave with respect to anaphoric relations. As Asher (1987) notes, it should be able to account for the fact that indefinites used under belief attributions (and indefinites used in the main context) can be picked up by anaphoric expressions the scope of *doubt*, but indefinites used under the scope of *doubt* can in general not figure as the syntactic antecedent of an anaphoric expression:

(1) John believes that *a woman* broke into his apartment.
He doubts that *she* left some fingerprints.
(2) *John doubts that *a woman* will marry him. He believes *she* will be unhappy.
(3) *John doubts that *a woman* will marry him. He doubts *she* will be happy.

Sometimes, however, indefinites used under the scope of *doubt* can be picked up by an anaphoric expression used under belief:

(4) John doesn't doubt that *a woman* broke into his apartment.
He believes that *her* perfume was unmistakably Channel No. 5. (Asher, 1987)

These data concerning the logical and anaphoric behaviour suggest that *a doubts that A* should be analysed as *a doesn't believe that A*, but in such a way that the embedded sentence can introduce properties.

$$[[\text{Doubt}^{P}(a, A)]](S) = \{\langle g',h,w\rangle \mid \exists g,h: \langle g,h,w\rangle \in S \;\&\; \sim\forall w' \in K(\mathbf{a},w): \exists g'',h': \langle g'',h',w'\rangle \in [[A]](\{\langle k,l,w''\rangle \mid k = g \;\&\; l = h \;\&\; w'' \in g(p)\}) \;\&\; \exists h'',w'':\langle g',h'',w''\rangle \in [[A]](\{\langle k,l,w''\rangle \mid k = g \;\&\; l = h \;\&\; w'' \in W\})\}$$

According to this interpretation rule only descriptive pronouns can 'refer back' to indefinites used under the scope of *doubt that*, and as we know, this is only allowed if the descriptive pronoun is interpreted with respect to a context in which the relevant property determines a singleton set in all worlds of this context. I think this explains (1) until (4). With respect to the logic, according to the above interpretation rules, the following inference (Asher, 1987) is predicted to be valid:

(5a) Fred doubts that either Mary or Alfred went to school.
(5b) So Fred doubts that Mary went to school and he doubts that Alfred went to school.

More in general, it is predicted that *downward entailment* is valid, just as wanted.[345]

6.3 Desire

6.3.1 A Hintikka-style analysis

Just like for the analysis of *doubt that*, to determine what the right analysis should be for verbs of desire (*want, wish, hope, be glad, need, look for, intend, try*) we will look both at inferential patterns and anaphoric relationships.

The first thing we want to account for is that intuitively, what one desires is connected with what one believes. One way to make sense of this intuition is to say that a sentence like *John wants A* is true iff all most desirable worlds compatible with what John believes make *A* true. This then looks very much like Hintikka's (1962) classical analysis of knowledge and belief. He assumed, like we did in chapter 1, that belief and knowledge states should be modelled by sets of possible worlds. Just like a belief state that is modelled by a set of possible worlds can be defined in terms of the propositions believed, also a desire state can be determined in terms of the propositions desired. For the analysis I will assume that for each agent and each world there is a set of propositions desirable for the agent. It seems reasonable to assume that the set of propositions desired might be mutually inconsistent, but it determines an ordering via a method used by Van Fraassen (1972) and Kratzer (1981). Thus, let G(j,w) be the set of propositions that John finds desirable in *w*. Then we say that *u* is at least as desirable as *v* with respect to G(j,w), $u \leq_{G(j,w)} v$, iff $\{A \in G(j,w) \mid v \in A\} \subseteq \{A \in G(j,w) \mid u \in A\}$.[346] World *u* can now be said to be strictly desirable to *v*

[260] Although the above interpretation-rule for *doubt-that* gets the above inferences right, it is doubtful whether it is simply the same as *doesn't believe that*. As Asher (1987) notes, it seems that verbs like *doubt that* requires a background justification for this doubt. I will come back to this later.

[346] In this way, $\leq_{G(j,w)}$ determines a partial ordering, but not a total one. Not all worlds have to be connected with each other.

with respect to G(j,w), u $<_{G(j,w)}$ v, iff u $\leq_{G(j,w)}$ v, but not v $\leq_{G(j,w)}$ u. On the basis of this ordering relation, we can define a function, Bul(j,w,X), that gives us the set of most desirable worlds in X with respect to the ordering relation determined by G(j,w):

$$Bul(j,w,X) := \{w' \in X | \sim \exists w'' \in X: w'' <_{G(j,w)} w'\}$$

On the basis of this function, we can now state the interpretation rule for factual desire sentences:

$$[[Want(j, A)]](I) = \{w \in I | Bul(j,w,K(j,w)) \subseteq [[A]](K(j,w))\}$$[347]

Although this interpretation rule is simple, it has a number of problems. First, the above interpretation rule can only account for desire sentences where the complement of the desire verb is consistent with what is believed. But we also make desire attributions where this is not the case, in those cases we typically use the verb *wish*. Second, the interpretation rule predicts that if one believes that *A*, one automatically also wants that *A*, which is a surprising prediction. Third, the rule predicts that desires are closed under logical implication, which does not seem to reflect the facts. Fourth, it is predicted that conjunction introduction is valid, although this doesn't seem to correspond with the data. The fact that it is possible that John wants to be with his wife, and that he wants to be with his mistress, although for obvious reasons he doesn't want to be with both, suggests that the set of desires that one has need not be consistent, and thus that for the analysis of desire attributions we should not only look at the most desirable worlds consistent with what is believed.

The problem that desires are predicted to be closed under logical implication, and that it is predicted that everything that is believed is also desired, hang closely together. The reason is that the proposition expressed by the complement of a desire attribution can entail something that is already believed. But this already believed proposition need not be desired by itself. If Robert believes that a murder has been committed, he wants to know who committed the murder. But knowing who committed the murder entails that the murder is committed. Still, it is not the case that Robert is happy that a murder has been committed. In the same vein, if Irene believes that she has to teach next semester, and wants to teach on Thursday, it can still be the case that Irene prefers not to teach at all next semester. And as we have seen already in chapter 4, if Rob believes that Mary has a vacuum cleaner that is broken and hopes that Sue knows that Mary's vacuum cleaner is broken, it does not follow that Rob also wants that Mary's vacuum cleaner is broken, although this is predicted by the above rules.

6.3.2 Desire as ceteris paribus preference

Asher (1987) and Heim (1992) concluded on the basis of data like these that we should not interpret desire attributions in a Hintikka style. Heim (1992) proposes that an attribution like *John wants A* is true iff John prefers *A* above *~A*. In this way, she gets rid of the closure condition for rational desires. Moreover, she assumes a *ceteris paribus* analysis of preference: *A* is preferred to *B*, if for every situation compatible with what is believed its closest world in which *A* but not *B* is true is preferred to its most similar world where *B* but not *A* is true.[348] If we assume that *f* is a similarity function as defined in chapter 5, and that in *w* John prefers proposition X to proposition Y, $X \leq_{j,w} Y$, iff $\forall w' \in X: \forall w'' Y: w' \leq_{G(j,w)} w''$ & $(Y = \emptyset \Rightarrow X \leq Y)$, Heim's interpretation rule for *want that* goes as follows:

---

[347] This was in fact the interpretation rule for *want that* proposed by Heim (ms).

[348] For a defence of this ceteris paribus analysis of preference, see Von Wright (1963) and especially Hansson (1989).

$[[\text{Want}(j, A)]](I) = \{w \in I \mid \forall w' \in K(j,w): f_{w'}([[A]](K(j,w))) \leq_{j,w} f_{w'}([[\sim A]](K(j,w)))\}$

If the analysis of *ceteris paribus* preference is to be preferred above the earlier assumed Hintikka-like analysis of preference, Heim's interpretation rule for desire attributions is arguably to be preferred above our earlier discussed analysis of desire attributions in exactly the same sense. The claimed advantages are that not only the most preferable worlds in a set count, and that rational desires are not predicted to be closed under logical implication.

Although the analysis of preference implicitly used by Heim (1992) verifies the principle that if A is at least as preferable to B, A is also at least as preferable to $A \vee B$, which in turn is at least as preferable to B, it still doesn't verify the stronger principle that says that if A is strictly preferred to B, and A and B are both compatible with what is believed, A is also strictly preferred to $A \vee B$, which in turn is also strictly preferred to B. This principle comes out valid if we have a logic that gives $A \vee B$ a preference value somewhere *in between* the preference values of A and B. That this is needed can be shown by the following example due to Rescher (1967):

Suppose we have four relevant worlds, $\{w_1,w_2,w_3,w_4\}$, where the propositions A and B differ in truth-value such that A is true in w_1 and w_2 and false in the other worlds, while the opposite is true for B. Suppose now that the ordering relation between possible worlds is such that w_1 is strongly preferred to w_4 which is just a bit better than w_2, which in turn is strongly preferred to w_3. Suppose now that except for A and B, w_1 is closest to w_3, and w_2 closest to w_4. In this situation, the *ceteris paribus* preference analysis would predict that A is not preferred to B and so that A is not wanted, which is, I think, clearly wrong. For instance, let us consider the preference ordering of a German general who wants to know whether he should attack France via Belgium, A, or directly via the German-French border, B. The worlds w_1 and w_3 are very close to each other because in those worlds the French only expect a German attack directly via the German-French border. In worlds w_2 and w_4, on the other hand, the French are well prepared for a German attack both via Belgium and via the direct border. If A is true in w_1 and w_2, and B in w_3 and w_4, clearly w_1 is strongly preferred above w_2, and w_4 is strongly preferred above w_3. Obviously, w_1 is strongly preferred to w_3: w_1 means victory and w_3 means defeat, because it is assumed that the French army is equally good as the German army. It also seems reasonable to assume that if the French are prepared for an attack at both places, it is better to attack directly via the German-French border, because of limiting transport problems. So, w_4 looks a bit better to the German general than w_2. But although there is a B-world, w_4, that is strictly preferred to an A-world, w_2, the German general is advised to attack the French via Belgium, and have the chance of an easy victory in battle. But according to the *ceteris paribus* analysis of preference, we should not advice the general to go via Belgium.

How can we get rid of this problem? The answer is simple: By using a more fine grained preference logic. The most suitable logic for our purposes seems to be (a variant of) Jeffrey's (1965) preference theory, to which I will now turn.

6.3.3 Desire as quantitative preference

Nice from our point of view is that Jeffrey's theory of preference, in distinction to some other quantitative preference logics, is compatible with the Boolean analysis of the connectives common in semantics. In the following discussion we will define probability and desirability measures on the algebraic structure given by the powerset of a given set of possible worlds.

Structure = $\langle W, \wp(W), P, d\rangle$

W a set of possible worlds

d $W \rightarrow \mathbb{R}$ (a function which assigns to each possible world its desirability according to the agent)

P $W \rightarrow [0,1]$ (a function which assigns to each possible world its probability according to the agent)

The *probability* of proposition *A*, P(A), is simply the sum of the probabilities of the cases (worlds) in which it is true, $P(A) = \Sigma_{w \in A} P(w)$. The *desirability* of a proposition *A*, d(A), is a weighted average of the desirabilities of the worlds in which it is true, where the weights are proportional to the probabilities of the worlds,

$$d_j(A) = \frac{\sum_{w \in A} P_j(w).d_j(w)}{\sum_{w \in A} P_j(w)} = \frac{1}{P_j(A)} \sum_{w \in A} P_j(w) d_j(w).^{349}$$

Given Jeffrey's preference theory, the simplest idea would be to say that a desire attribution *John desires that A* is true if the desirability for John of the embedded clause is larger than the desirability of a tautology:

$$\text{Desire}(j, A) = 1 \text{ iff } \frac{1}{P_j(A)} \sum_{w \in A} P_j(w) d_j(w) > \sum_{w \in T} P_j(w) d_j(w) \text{ iff } d_j(A) > d_j(T)$$

It is easily seen that this doesn't predict desires to be closed under logical consequence, that it doesn't make conjunction introduction valid anymore, and that it can account for Rescher's problem.[350] In distinction to the analysis of bulletic predicates by Heim, it doesn't make use of the *ceteris paribus* condition, but in this case it's not needed to get a very weak system.

Jeffrey's system predicts that desires are not closed under logical consequence, and it predicts that if Desire(j, A) is true, and John prefers B to A, also Desire(j, B) is true. This sounds all very nice, but the proposal gets us into one serious problem. It cannot make sense of desires that are already believed to be true or believed to be false. This is due to Jeffrey's use of conditionalisation of absolute (classical) probability functions. as a result, we will never look outside of that what we believe. But now that we know what the problem is, the solution is obvious. Use Popper functions that we discussed in chapter 5.[351] Remembering that the change function C corresponded with imaging, all we have to do is to make out of imaging, *preservative* imaging.

Given a probability function P and any proposition A, there is a probability function $P^{i}P_A$ such that, for any world w':[352]

---

[349] The given formulae are for simplicity based on the assumption that there are only finitely many possible worlds. If there are infinitely many possible worlds in which a certain proposition is true, every possible world in this set has probability 0. It is important, however, that Jeffrey's theory is not based on one of these assumption. Desirability can of course also be calculated for continuous probability functions, but to do that we have to make use of intervals and integrals.

[350] Rescher's (1967) logic of preference can also handle those problems, but that is no big surprise. Rescher's logic is only a special case of Jeffrey's system. For Rescher all possible worlds have equal probability. It is clear that this makes Rescher's logic unsuitable for decisions under uncertainty.

[351] Heim's *ceteris paribus* analysis of preference faces of course a similar problem, and can be solved by a similar solution: use preservative revision instead of [[ ]].

[352] For simplicity we will assume the uniqueness assumption, but as we saw in a footnote of § 5.6, this is not necessary.

$$P^{ip}{}_A(w') = \Sigma_{w \in K} P(w) \times \begin{cases} 1 \text{ if } f^K{}_w(A) \text{ is } w' \\ 0 \text{ otherwise} \end{cases}$$

Desire attributions can now be interpreted in the expected way:

$$\text{Desire}(j, A) = 1 \text{ iff } \Sigma_{w'} P_j{}^{ip}{}_A(w') \times d_j(w') > \Sigma_w P_j(w) \times d_j(w)$$

Let's consider our model again with four worlds, where w_1 and w_3 are most similar to each other, and the same holds for w_2 and w_4. Let us also assume that $A = \{w_1, w_2\}$, and $B = \sim A = \{w_3, w_4\}$, and that all four worlds are equally likely true. In that case, the *ceteris paribus* analysis of preference demands that for A to be desired, both w_1 must be preferred to w_3, and that w_2 must preferred to w_4. Jeffrey's preference theory, on the other hand, only demands that if we can give a cardinal valuation to the four worlds, that the average valuation of w_1 and w_2 is higher than the average valuation of w_3 and w_4. As this example illustrates, the quantitative approach *weakens* Heim's qualitative approach. In the quantitative approach, we don't compare possible worlds that are most similar to each other, but instead we compare whole information states. Whether the weakening is in general preferred to Heim's strong notion of preference, I don't know, but I think that it's preferred in at least some cases.

6.3.4 A conditional analysis of desires

Until now we have discussed three kinds of analyses of desire attributions. The first was based on a classical Hintikka-like all_or_nothing analysis of preference, the second was based on a *ceteris paribus* analysis of preference, and the third on a quantitative notion of preference. In this section I will discuss yet another analysis of preference, and this analysis will be related later to a popular analysis of permission sentences.

Asher (1987) observed that desire attributions normally obey disjunction elimination, and Zimmermann (ms.) observed that indefinites in the scope of verbs of desire are normally interpreted "arbitrarily". Thus, we can normally infer (6b) from (6a), and (7a) is normally interpreted as something like (7b):

(6a) Alexis hopes that she will have chicken or fish for dinner.
(6b) So she hopes that she will have chicken for dinner.
(7a) John wants to catch a fish.
(7b) John wants to catch an arbitrary fish, any fish will do.

These facts are surprising for any of the above proposals. They can, however, be accounted for if we assume that desire attributions should be understood as implicit conditionals. Thus, if *John wants that A* means something like "If A is the case, John will be satisfied". Disjunction elimination now follows immediately, but unfortunately, also the more general *downward entailment* is predicted to be valid. It is predicted that if John wants A, and B entails A, it follows that John wants B, too. But this is obviously a wrong prediction: I want to have a holiday this summer, but do not want a holiday and bad weather. Still, the conditional interpretation of desire attributions can be rescued, if this conditional is not treated as an indicative conditional, but as a *subjunctive* conditional instead. To make sense of this, we can assume that K(j,w) represents no longer the set of futures consistent with what John believes in *w*, but the possible ways the world might be *at this moment* according to John in *w*. Thus, if we want to look at the future, we have to use already the more general revision rule. I will assume that if somebody wants *A*, he has a desire about the future and so does not believe it yet. Desire attributions can now be analysed in terms of revision as follows:

$[[\text{Desire}(j, A)]](S) = \{w \in S \mid C''_{K(j,w)}(A) \subseteq \text{Bul}(j,w,W)\}$[353]

Thus, John wants A in *w* is true iff K(j,w) revised by A is a subset of the set of John's absolute favourites among the worlds. Note that according to the above rule, neither upward entailment, nor downward entailment is valid. Moreover, disjunction elimination is allowed, but only if the complements of both disjuncts are equally strongly entrenched.[354] This seems exactly what we need. Normally disjunction elimination is valid, and normally indefinites get the arbitrary interpretation, but this is not always the case:[355]

(8) John wants a beer, but not a warm one.

6.3.5 Rational desires and relevant alternatives

Still, a counterexample like (8) to the non-arbitrarily interpretation of the indefinite has intuitively nothing to do with *epistemic* entrenchment. This suggests that the ordering relation by which we determine the relevant change function should not be induced by epistemic entrenchment, but by desirability, instead. This suggest that we should use the following interpretation rule:

$[[\text{Desire}(j, A)]](S) = \{w \in S \mid \text{Bul}(j, w, A) \subseteq \text{Bul}(j, w, W)\}$

Before we come to the problematic aspect of this interpretation rule, let us first look at the good part. First, it allows only for disjunction elimination, and thus for the arbitrarily interpretation, if the disjuncts are equally desirable. Second, it can account for the fact why sequences like (9) are out:

(9) John wants a cool beer, but he doesn't want a beer.

The reason is, according to this approach, that desires are closed under logical implication. It can easily be checked that according to the above interpretation rule the sentence *John wants A* is true in *w* for any $A \neq \emptyset$ iff $A \cap \text{Bul}(j,w,W) \neq \emptyset$, because if $A \neq \emptyset$, then also $\text{Bul}(j,w,A) \neq \emptyset$. It then immediately follows that if $A \subseteq B$, also $B \cap \text{Bul}(j,w,W) \neq \emptyset$, and thus John wants B, too.

Note that although according to this rule rational desires are predicted to be closed under logical consequence, the rule doesn't predict that desires have to be mutually consistent with each other. That is, it is not predicted that conjunction introduction is valid for rational desires. And, as the wife and mistress problem discussed in § 6.3.1 suggested, that is the way it should be.

But in the same section we saw that the closure condition for rational desires is also problematic. We have seen above some examples that show that if John wants A, and B follows from A and is already believed by John, it doesn't have to be the case that he also wants B. As far as I can see, the only problems for the closure under entailment for rational

---

[353] where Bul(j, w, W) is defined as is § 6.3.1. The form of this interpretation rule was actually proposed by Price (1989) in his defence of the Desire-as-Belief thesis. An alternative formulation would be to use imaging, defined in terms of a fixed selection function *f*. As we have seen in chapter 5, in that case we should not expect that it always holds that if A is consistent with K, $C_K(A) = K \cap A$.

[354] If the revision function C'' obeys all 8 Gärdenfors postulates, $C''_K(A \vee B) = C''_K(A) \cup C''_K(B)$, if ~A and ~B are equally strong entrenched in K.

[355] The following counterexample to disjunction elimination, due to Ede Zimmermann (personal communication), suggests that decision theory is relevant for the analysis of at least some desire attributions after all : if Alexis thinks that there is a tiny chance of getting chicken and a good chance of getting fish (both of which she prefers to anything else she considers possible), then (6a) seems to be OK, but (6b) isn't.

desires that cannot be solved in the same kind of way as the wife and mistress problem was solved are of this nature. But then we should conclude that the main problem with the above interpretation rule is that desires are no longer related to beliefs.

If we relate desires to belief we can solve the latter problem, too, if we follow Stalnaker (1984) and relativise the consequent relation for rational wants:

> Some propositions which are entailed by propositions that one wants to be true in this sense are also entailed by the relevant alternatives. It is not that I want these propositions to be true - it is just that I accept that they will be true whether I get what I want or not. Given that there was a murder, I would rather know who committed it than not know. The question whether or not I look with favour on the fact that there was a murder - whether I am glad that it happened or wish it had not - does not arise in that context. To raise *that* question, one needs to expand the set of relevant alternatives, to compare the actual situation with possible situations in which the murder never took place.
> The qualified consequence condition for rational wants motivated by these considerations is this: the propositions one wants to be true (relative to a set of relevant possibilities) includes all the consequences of any proposition one wants to be true *which distinguish between the relevant alternatives*. (Stalnaker, 1984, pp. 89-90)

It is clear that to determine whether *A* is wanted or not, we should consider worlds, as close as possible to the belief worlds, where *A* is the case, and such worlds where ~A is the case.

To formally account for this suggestion, it seems we should not compare the best of all A-worlds with the best of all worlds, but compare the best of all A-within a contextually given set with the best of all worlds in this contextually given set. This contextually given set must then be such that it contains some A-worlds and some ~A-worlds. Moreover, for the analysis of *want that* it seems that normally this contextually given set is the set of worlds compatible with what the agent believes.[356] As a result, we can interpret desire attributions of the form *a want A* in the following way (where K(j,w) represents the beliefs about past, present and future of John in w):

$$[[\text{Want}(j, A)]](S) = \{w \in S \mid \text{Bul}(j, w, [[A]](K(j,w))) \subseteq \text{Bul}(j, w, K(j,w))\},^{357}$$

and presuppose that A is true in some but not all worlds of K(j,w) for all $w \in S$. This rule is the same as the interpretation rule for *desire* above, except that it is relativised to the belief state of the agent. By means of revision, we can also easily implement the suggestion of Stalnaker to account for desire attributions made by verbs like *wish that* and *be glad that*:

$$[[\text{Wish}(j, A)]](S) = \{w \in S \mid \text{Bul}(j, w, [[A]](K(j,w) \cup C''_{K(j,w)}(A))) \subseteq \text{Bul}(j, w, K(j,w) \cup C''_{K(j,w)}(A))\}$$

$$[[\text{Glad}(j, A)]](S) = \{w \in S \mid \text{Bul}(j, w, [[A]](K(j,w) \cup C''_{K(j,w)}(\sim A))) \subseteq$$

---

[356] Normally, because (i) in some *want* attributions the context of interpretation for the embedded clause needs to be a *superset* of the belief state, as for Heim's (1992) example *(John hired a baby-sitter because) he wants to go to the movie tonight*, and (ii) sometimes the context of interpretation should be a *subset* of the belief state, as for desire attributions conditionally dependent on other desire attributions: *John's father hopes that his son never smoked before, and hopes that he just started smoking*. See also Geurts (1995).

[357] After the earlier given interpretation rule of *desire* and the quote from Stalnaker, Ede Zimmermann still had to point out to me that this rule was the natural consequence. I suggested, instead, a different interpretation rule. Heim (ms) proposes an analysis that looks very similar:
$[[\text{Want}(j, A)]](S) = \{w \in S \mid \text{Bul}(j, w, [[A]](K(j,w))) = \text{Bul}(j, w, K(j,w))\}$.
The distinction is that Heim proposes in this way a Hintikka-like analysis of *want that*, while I don't. Heim (ms) requires that the best K-worlds are A-worlds, while I require that the best A-worlds in K are all among the best K-worlds.

$$\text{Bul}(j, w, K(j,w) \cup C''_{K(j,w)}(\sim A))\},^{358}$$

where it is presupposed that A is incompatible with K(j,w) for *wish that*, and that A is entailed by K(j,w) for *be glad that*.

Note that if all of A, ~A, B and ~B are consistent with K(j,w), and if John wants A, it follows that John also wants B, if B follows from A. If we then assume that K(j,w) represents again the possible futures compatible with what John believes in *w*, we can account for the fact that factual desire attributions allow for existential weakening, as exemplified by the contradictority of (9).[359]

On the other hand, disjunction elimination and the arbitrarily interpretation of indefinites used in desire attributions are not valid according to the above interpretation rule. From *John wants that A or B* I can only conclude that John also wants A, if A is at least as desirable for John as B. Similarly, from *John wants an apple*, I can only conclude that John wants a green apple, if eating green apples is at least as desirable for John as eating apples of any other colour. And this is confirmed by (8).

Until now I have assumed that all verbs of desire should be analysed in the same way. Thus that the emotive cognitive attitude *hope* should be analysed in the same way as a pro-attitude like *intend*. Intuitively, however, there are two differences between *intend* and *hope*: (i) whereas what you intend has typically something to do with your own activities, *hopes* are not so closely related with actions of the agent himself, and (ii) whereas *intend* is necessary future oriented, *hope* need not be, as in *I hope he survived the operation*. For *intend* it is normal to take as complement *to* infinities that seem to designate abilities, but as observed by Portner (1992, ch. 4) the verb *hope* takes also that-clauses as complements. I don't want to suggest that the two verbs should be analysed in a completely different way, but it might be the case that we use two different concepts of desire, one concept about futures that the agent can influence himself, and one that is about circumstances he cannot influence. Moreover, that these two concepts are typically expressed by the words *intend* and *hope*, respectively. One option to 'explain' this all is to say that the truth conditions for these constructions are identical and should be analysed as before, but that appropriateness conditions for asserting such sentences differ. For *hope* it should be the case that both the embedded clause and its negation should be consistent with what is believed about the present by the agent and with the global context, but this need not be the case for *intend*. Maybe the intuitive difference between the two concepts can be accounted for in this way, but maybe we should take the notion of *action* more serious.

But once we assume that desires should be treated as possibility operators, there might be a more obvious way to account for the difference between desire and intention: just analyse intention as the corresponding necessary operator of desire.

$$[[\text{Intend}(j, A)]](S) \quad = \quad \{w \in S \mid \text{Bul}(j, w, K(j,w)) \subseteq A\}$$

In this way it is predicted that rational intentions are not only closed under logical consequence, but also that we cannot have incompatible intentions. But what if the best worlds in K(j,w), Bul(j, w, K(j,w)), are all B worlds, although if John would slightly revise his beliefs by a new proposition D to $C''_{K(j,w)}(D)$, B would not even be compatible

---

[358] These interpretation rules are very similar to the interpretation rules given by Heim (1992).

[359] Graham Katz gave a talk at the Sinn und Bedeutung conference at Tübingen, 1996, where he argued for a Hintikka-like analysis of desire attributions. For me the most convincing argument was that (9) really seems contradictory. Only after hearing this talk I realised that the remarks of Stalnaker (1984, p. 90) should probably be interpreted in the way as suggested above.

with Bul(j, w, C"$_{K(j,w)}$(D)), would we still say that John intends also B? I think not.[360] I want to suggest tentatively that *John intends A* is true iff B is true in all of the best of John's belief worlds, and almost no amount of further information would change that.[361] To implement this suggestion, let us assume that S is a set of contextually given propositions which contains always the tautological proposition. Then I propose:

$$[[\text{Intend}(j, A)]](S) = \{w \in S \mid \forall B \in S: \text{Bul}(j, w, C"_{K(j,w)}(B)) \subseteq A\}.$$

In this way it is predicted that if John intends A, he also wants it, if ~A is compatible with what he believes.

6.3.6 Desires and anaphora

Although I believe that the last interpretation rule can account for the inferential patterns, it cannot account for the possible anaphoric relations across attitude attributions where the antecedent stands in the scope of a desire verb, like in (10) and (11):

(10) Sue wants to marry *a Swede*, and she wants a child of *him*.
(11) John wants to catch *a fish*, and he wants to eat *it* afterwards.

On their most natural interpretations, the indefinites of the first clauses are not specifically used by the speaker. They are neither intended to refer to a specific Swede or fish, nor to a specific belief object of the agent. As a consequence, according to the theory of anaphora I defended in chapters 2 and 3 of this dissertation, the pronouns occurring in the second clauses can only be used as descriptive pronouns. But for a descriptive pronoun to be used appropriately, it has to be presupposed that the relevant property denotes a singleton set in each possibility of the relevant context of interpretation. There is no reason why this should be true for (11), however, if we assume the above interpretation rule for desire attributions.

What we should do is make the context of interpretation for the second desire attributions not the belief state of the agent itself, but this belief state 'revised' by the complement of the first desire attribution. But how should we think of this revision such that we can account for uniqueness? I want to propose that the relevant revision for *want + to-infinitive* has the effect that we don't look at what the agent believes about the present, but about future states of worlds that are compatible with what the agent believes. I want to propose that we should add to the possibilities an extra element, namely a time-point, and that sentences are represented with an extra reference point marker that will always be interpreted as the time of the possibility.[362] For instance, the sentence *John catches a fish* will be represented by $\exists x[\text{Fish}(x) \wedge \text{Catch}(j,x,r)]$, and for all possibilities $\langle g,h,w,t\rangle$ it will be the case that $r \in \text{dom}(h)$ and $h(r) = t$. Important is that the properties introduced by indefinites are now no longer functions from worlds to sets of individuals, but functions from world-time pairs to sets of individuals. This relativation to time-points has the effect that for action-verbs like *catch* uniqueness of the property associated with *Fish caught by John* is reached much easier. To account for desire attributions of the form *want that* I will make two extra assumptions: First, with desire attributions we don't compare worlds with each other, but world-time pairs; Second, with a desire attribution of the form *want that* we push the reference point of each relevant possibility forwards. The 'proposition' introduced by such a desire attribution is then a set of world-time pairs that represents possible future states of the worlds compatible with what the agent believes about the present. As a result I will interpret these desire attributions as follows:

---

[360] Bratman (1987) convincingly argues that what one intends is a subset of what one chooses. Expected side effects of what one chooses need not always be intended.
[361] Compare this with the analysis of evidential verbs in § 6.5.1.
[362] For a much more detailed treatment of tense in intensional contexts, see Abusch (ms.), and Heim (1994).

$[[\text{Want}^{p}_{q}(a, A)]](S) = \{<g',h',w,t>| \exists g,h,h'': <g,h,w,t> \in S:$

$\text{Bul}(\mathbf{a},w, \text{WT}([[A]]\,(\{<k,l,w'',t''>| k = g \,\&\, l = h \,\&\, \exists t':$

$<w'',t'> \in g(p) \,\&\, t' < t''\}) \subseteq \text{Bul}(\mathbf{a},w,\{<w',t''>| \exists t': <w',t'> \in$

$g(p) \,\&\, t' < t'') \,\&\, h''[q]h' \,\&\, h'(q) = \text{WT}([[A]](\{<k,l,w'',t''>| k = g \,\&$

$l = h \,\&\, \exists t': <w'',t'> \in g(p) \,\&\, t' < t'' \,\&\, \exists k,w'',t'': <g',k,w'',> \in$

$[[A]](\{<k,l,w'',t''>| k = g \,\&\, l = h \,\&\, \exists t': <w'',t'> \in g(p) \,\&\, t' < t''\})\}$

Note that by this interpretation rule the sentence *John wants to catch a fish* will normally have an arbitrarily interpretation, and that the 'proposition' introduced by the embedded clause will be the set of future world-time points (with respect to the agent's *now*) where John catches a fish in that world at that time. Note also that once we assume that desire sentences can introduce 'propositions', we can also account for the intuition that the context of interpretation of a belief sentence can be dependent on a foregoing desire sentence. And this is needed, too, as can be illustrated by the following example.

(12) Slave John wants to catch *a fish*,
although he believes that his master will eat *it* afterwards.

6.4 Epistemic attitudes analysed in terms of plausibility

6.4.1 Plausibility

The entrenchment relation used for the analysis of belief revision gives rise to a notion of *plausibility*. First, it gives rise to a plausibility grading of the possible worlds. On the basis of an information state K and a set of propositions S that potentially determines similarity, we can define a function, k, that represents the same information: $|S^{w}{}_{u}C''_{K}(A)|$, for any u in $C''_{K}(A)$. The measure k(w/A) represents the plausibility of w after revising the information state K with A. The idea is that k(w/A) is the number of propositions decided by $C''_{K}(A)$ that potentially determine similarity on which w and any arbitrary element of $C''_{K}(A)$ differ in truth-value. The higher k(w/A) is, the less plausible the agent in belief state k would find world w after he would revise his belief state by A. In terms of k(w/A) we can define $k(B/A) := \min\{k(w/A): w \in B\}$. The measure k(B/A) represents the degree of disbelief in B given that A is true. If k(B/A) = 0, this means that B is consistent with the belief state resulted after revision of our current belief state with A. Given the definition of k(B/A), we can define the plausibility of other complex propositions after conditionalising on A. Thus $k(B \vee C/A) = \min\{k(B/A), k(C/A)\}$, and $k(B \wedge C/A) = k(B/A) + k(C/A \wedge B)$. This looks a lot like probability, and indeed, reading '+' for 'min', and 'x' for '+', it satisfies the main constraints on Popper functions. It is also easy to see that once we conditionalise on T, a tautology, the plausibility function k satisfies the main laws of standard probability calculus. For those who have seen Spohn's (1987) it is obvious that the above plausibility functions are a simplistic variant of his *ordinal conditional functions*. Let us abbreviate k(A/T) by k(A). Then we can follow Spohn in saying that A is accepted in k, (or in K) iff $k(\sim A) > 0$. We noted that we can define k in terms of K (+ a set of propositions S), but we can also go the other way round, $K := \cap\{A \subseteq W| k(\sim A) > 0\}$. What k measures is potential surprise. In general, we can say that A is believed to be more plausible than B, A > B, iff $k(\sim A) > k(\sim B)$ or $k(A) < k(B)$. Given the close relation between our entrenchment relation and Spohn's ordinal conditional functions, it should not come as a big surprise that our entrenchment relation satisfies the five Gärdenfors postulates (1988, pp.88-91) for entrenchment: For all A, B, C, $K \subseteq W$:

(EE1) if $A \leq_K B$ and $B \leq_K C$, then $A \leq_K C$,

(EE2) if $A \subseteq B$, then $A \leq_K B$,

(EE3) for all $A, B \supseteq K$, $A \leq_K A \cap B$ or $B \leq_K A \cap B$,

(EE4) if $K \neq \emptyset$, for all $B \supseteq K$, $K \not\subseteq A$ iff $A \leq_K B$, and

(EE5) if $B \leq_K A$ for all B, then $A = W$.

The reason is of course that the following equation holds: $A \leq_K B$ iff $k(\sim A) \leq k(\sim B)$, and that it is well known (see Gärdenfors, 1988, §4.6) that Spohn's ordinal conditional functions generates an entrenchment relation that satisfies (EE1)-(EE4) and that (EE5) follows from possible world semantics.

Let us now assume that a belief state should not be represented by a set of possible worlds, but rather by a plausibility function, whether such a function can be derived from such a set plus a set that potentially determines similarity or not. It seems reasonable to assume that once we have such a richer representation of a belief state, we can account for more attitudes in terms of belief states than would be possible without such a representation. Indeed, this is what I will assume.

6.4.2 Evidential verbs

It seems reasonable that verbs like *be certain*, *be sure*, *be convinced* and the future looking *expect*, and *predict* should be analysed as *believe* + some extra condition. The reason is that from *a is sure that A* we can conclude that a believes that A. What should this extra condition be? Following Asher (1987), it should at least guarantee the principles of *belief inference,* (B), *simplification*, (S), *conjunction introduction*, (I&), and *upward entailment*, (UE), for all these kind of verbs (where α is the attitude verb):

B:	a α that A	$\Rightarrow$	a believes that A
S:	a α that $(A \wedge B)$	$\Rightarrow$	a α that A
I&:	a α that A & a α that B	$\Rightarrow$	a α that $(A \wedge B)$
UE:	a α that A & $A \subseteq B$	$\Rightarrow$	a α that B

The extra condition for these verbs is *evidential* in nature, it should be some kind of *justification condition*. The simplest way to make these inferences come out right, is simply to assume a new accessibility function. In this way principles (S), (I&) and (UE) follow immediately. To account for (B), this new evidential accessibility function assigns to each world *w* a set of worlds that is a superset of K(a,w), the doxastically accessible world. Although this kind of rule will do to account for the above inferences, it is preferred to account for these principles by using primitives we use already, or by primitives that are also useful for the analysis of other attitudes. I propose to account for the inferences by using the extra *inductive* information represented in a belief state if we take the notion of *epistemic entrenchment* seriously. If α is an evidential verb, *a* α *that A* is true only if *a* believes that A, and A is highly entrenched in a's belief state. In other words, it should be the case that A is believed, and that ~A is very unplausible, or that A is very strongly believed. We have seen above that given a set of propositions that potentially determines similarity, a belief state K gives rise to an ordinal function, k, that measures implausibility, which in turn, via the *Shackle identity*, $f(A) = k(\sim A)$, gives rise to a belief function, *f*, that

measures plausibility or epistemic entrenchment.[363] Let us say that $f^{a,w}$ is the belief function associated with *a* in *w*. So my proposal comes down to the following:

$w \in [[a\ \alpha\ \text{that}\ A]](S)$ only if $f^{a,w}(A)$ is high

What *high* means is context dependent, but the number should be at least bigger than 0. Note that if $f^{a,w}(A) > 0$, then $k^{a,w}(\sim A) > 0$, and thus $K(a,w) \subseteq A$. The proposed account predicts at least that from the truth of *a* α *that A* we can infer that A is believed by *a*. Let us now see whether it also can account for the other inferences. Note first that simplification is a special case of upward entailment. Thus, if the above definition accounts for (UE), it also accounts for (S). We know already that if K is the belief state corresponding with k, and if we defined the entrenchment relation $\leq_K$ between propositions by $k(\sim A) \leq k(\sim B)$ iff $A \leq_K B$, then the entrenchment relation will satisfy the Gärdenfors postulates for (EE1)-(EE5). In particular it satisfies (EE2), if $A \subseteq B$, then $A \leq_K B$, and this is enough to guarantee that our interpretation rule for evidential attitudes accounts for (UE) and thus for (S). Conjunction introduction also follows from our interpretation rule. Note that by (EE3) if both A and B are believed, then either $A \leq_K A \cap B$, or $B \leq_K A \cap B$. Thus either $f^{a,w}(A) \leq f^{a,w}(A \wedge B)$ or $f^{a,w}(B) \leq f^{a,w}(A \wedge B)$. But if both $f^{a,w}(A)$ and $f^{a,w}(B)$ are high, then also $f^{a,w}(A \wedge B)$ must be high, and thus conjunction introduction, (I&), also holds for evidential attitude verbs. By the way I interpreted evidential attitudes, these attitudes have the properties of acceptance attitudes. An acceptance attitude is an attitude that can be modelled by an *acceptance state*. An acceptance state is a consistent set of proposition closed under conjunction and implication. By modelling propositions as sets of possible worlds, the intersection of an acceptance state gives rise to a set of possible worlds. How can we arrive at this set of worlds from the above interpretation rule of evidential attitudes? Above we have assumed that *a is certain that A* iff $f^{a,w}(A)$ is high. Let us now say that *n* is the minimum of the high numbers. Remember that via the Shackle identity, $f^{a,w}(A) = k^{a,w}(\sim A)$, and that $k(A) = \min\{k(w): w \in A\}$, where $k(w) = k(w/T)$ and T is a tautology. The evidential accessibility function, EVI, can now be determined in the following way: $EVI(a,w) = \{w' \in W: k^{a,w}(w') \leq n\}$. Suppose now that EVI(a,w) is the epistemic accessibility relation. Because $n > 0$, K(a,w) will be a subset of EVI(a,w), just like Hintikka (1962) demands to account for the inference from knowledge to belief. Note also that Spohn's ordinal conditional functions were developed to account for belief revision, and that Hintikka's prime intuition about knowledge was governed by stability under change of belief:

> It may be useful to remember that for us the primary sense of "I know that p" is the one in which it is roughly equivalent to "p, and no amount of further information would have made any difference to my saying so". (Hintikka, 1962, p. 52)

Of course, I don't demand that in general $w \in EVI(a,w)$, so I have not yet accounted for the factivity of knowledge.

---

[363] See Spohn (1987) who refers back to Shackle (1961). If you want to be more abstract, you can take *f*, or *k*, to be primitive and derive K from it. For the use of belief function's, entrenchment orderings and belief revision in non-monotonic logic, see for instance Gärdenfors and Makinson (1994). For the relation between entrenchment orderings and non-standard probability functions, see Spohn (1987), McGee (1994) and Pearl (1994).

6.4.3 Be surprised

Another attitude verb that is naturally interpreted in terms of an entrenchment relation is the verb *be surprised.* We will guide our investigation again by the principles it should obey. Contrary to evidential attitude verbs, *be surprised that* is not closed under implication. If John is surprised that it snows, he need not be surprised that it rains or snows. According to Asher (1987), *be surprised* is a negative factive and the interpretation rule for those verbs should obey *factivity,* (F), *belief inference,* (B), *negation,* (N), *weakened simplification,* (WNS), and *weakened downward entailment,* (WDE):

F:	a α that A	$\Rightarrow$	A is true
B:	a α that A	$\Rightarrow$	a believes that A
N:	a α that ~A	$\Rightarrow$	~(a α that A)
WNS:	a α that (A $\vee$ B) & a believes that A $\vee$ B	$\Rightarrow$	a α that A
WDE:	a α that A & B $\subseteq$ A & a believes that B	$\Rightarrow$	a α that B

That *be surprised that* should obey factivity is clear. Asher argues that the verb *be surprised that* should obey (B) because it is incoherent to say *Fred is surprised that John runs, but he doesn't believe it.* The inferences (F) and (B) should be *presuppositional* inferences, because the inferences are normally preserved under negation. Not only from a positive, but also from a negative sentence like *Mary was not surprised that John didn't get an A*, we can infer that John didn't get an A, and that Mary believed that John didn't get an A. I stated above that it is natural to interpret *be surprised that* in terms of epistemic entrenchment, because if A is believed, k(~A) is normally called *the surprise value* of A in artificial intelligence. The most natural interpretation for *a is surprised that A* would be: A is true, *a* believes A, and in the belief state before learning A, it was expected that ~A. Thus in this earlier belief state, k(A) was high. However, just like subjunctive conditionals should not always be interpreted with respect to a prior belief state (or objective state of affairs), it doesn't seem to be the case that for being surprised that A, I had to expect ~A in a prior belief state. First, according to some philosophical schools, the real way of being a philosopher is by being surprised about things you always took for granted. Second, suppose someone learns a mathematical theorem at a young age, and only after learning much more about mathematics he sees how deep the theorem really is, how surprising the truth of the theorem is given everything else he knows about mathematics at his current state. This suggests that we should not always look at an earlier belief state, but sometimes must be able to interpret *being surprised that* in terms only of the present belief state. To account for these latter cases, cases of surprised$_2$, my proposal would be the following:

a is surprised$_2$ that A = $\{w \in W | f^{a,w}(A) > 0$, but low$\}$

Note first that this interpretation rule for *being surprised that* does not predict that it is closed under implication. It is easy to imagine a situation where A is not strongly entrenched, but A $\vee$ B is so, because B is. If B is strongly entrenched, $f^{a,w}(B)$ will be high, and thus also $f^{a,w}(A \vee B)$ will be high. It follows that *being surprised that* will not be closed under logical implication. Like in the case of evidential predicates, it follows from being surprised that A that the agent also believes A, because $f^{a,w}(A) > 0$ iff $k^{a,w}(\sim A) > 0$. The principle of *negation* follows immediately from the definition. If *a* is surprised that ~A, then it should also be the case that ~A is believed by *a*, in which case *a* cannot be surprised that A. Now we have to show that *weakened negative simplification* and *weakened downward entailment* are obeyed. Note that (WNS) is a special case of (WDE), so it is enough to show that (WDE) holds. But this follows immediately from the interpretation rule. If $f^{a,w}(A) > 0$, but low, and B $\subseteq$ A, then via (EE2) and the Schackle identity, $f^{a,w}(B) \leq f^{a,w}(A)$. Because it is also assumed that B is believed, also $f^{a,w}(B) >$

0. It follows that if $f^{a,w}(A) > 0$, but low, also $f^{a,w}(B) > 0$, but low, and thus that it is also surprising that B.[364]

6.4.4 Doubt

As we saw in the beginning of this chapter, the interpretation of *doubt that* should be such that it is not closed under implication, but instead obeys *negative simplification*, (NS), and *downward entailment*, (DE). We have seen that this can be easily accounted for if we assume that we should interpret *a doubts that A* as *a doesn't believe that A*. Although this interpretation-rule for *doubt-that* gets the above inferences right, it is doubtful whether it simply means the same as *doesn't believe that*. Intuitively, I think, *doubt that A* means more something like *doesn't believe that A and his believe justifies not A*.[365] The question is, how should we interpret this justification-condition? Given our discussion above of evidential attitudes and of *being surprised that*, it will not be surprising that I propose to use the Shackle-Spohn plausibility functions again. The interpretation rules for *doubt* would then be:

a doubts that A = $\{w \in W | f^{a,w}(\sim A)$ is high$\}$

Note that now *a doubt that A* means the same as *If A were true, John would be surprised*. This interpretation rule for *doubt that* is very strong. It says that *a doubts that A* iff *a strongly believes that ~A*. Note that by this interpretation rule, downward entailment is still satisfied. Because If $B \subseteq A$, then $f^{a,w}(\sim A) \leq f^{a,w}(\sim B)$, it follows that if *a* doubts that A and $B \subseteq A$, *a* also doubts that B.

6.4.5 Plausibility versus probability

By the way I interpreted evidential attitude verbs and *doubt*, I have to assume that a lot of propositions are believed. Maybe too many propositions. Wouldn't it be easier and more appropriate to use probability instead of plausibility? Let $P_{a,w}$ be the probability function that represents the belief state of *a* in *w*, let *r* be any real number in [0,1], and let α be any evidential attitude. It looks as if it is more appropriate to analyse the different attitude verbs in the following way:

a α that A = $\{w' \in W | P_{a,w}(A) > r\}$
a doubt that A = $\{w' \in W | P_{a,w}(A) < r\}$

It is easily seen that for evidential attitudes, this also accounts for closure under implication and thus for simplification. If it is assumed that a believes that A if $P_{a,w}(A) > s$, where $0 < s \leq r$, we can also account for the belief inference (B). For *doubt that* it even accounts for all the demanded principles, addition, downward entailment and negative simplification. That seems to be pretty good, but there is a problem. Accounting for belief and evidential predicates by probabilistic predicts that the relevant attitude is not closed under conjunction, unless the relevant number is 1. I think that this is not only unwanted for belief, but also for the evidential predicates. This problem does not arise for the interpretation of *doubt that*, so there doesn't seem to be any good reason for not interpreting this predicate as

---

[364] Ede Zimmermann (personal communication) has given the following example meant as a counterexample to the proposed analysis: "Suppose Ede meets a friend in the street whom he had believed to be far away (or dead) and convinces himself that it is really her, then Ede would still be absolutely convinced yet at the same time surprised that she is there - at least before he learns the explanation." Is this a counterexample? Maybe not, if for his conviction and for his being surprised it are different sets of propositions that potentially determine similarity.

[365] That is, if we ignore the intuition that doubt seems to involve active thinking.

suggested. Moreover, the way we originally interpreted *doubt that* assumes that the set of believed propositions is very big. Still, it might be desirable to have a non-probabilistic account of all attitudes. Fortunately, it's not difficult to save the qualitative analysis by weakening the set K(a,w). Let us follow Kratzer (1981) and use *stereotypical backgrounds*. A stereotypical conversation background is a set of worlds representing what is normally the case. Let us assume that ST(a,w) represents that what *a* in *w* thinks is normally the case. Given a set of propositions, S, that potentially determines similarity, we can derive an ordinal function *ks* by S and K(a,w) $\cap$ ST(a,w) as we derived earlier the ordinal function *k* by S and K(a,w). The belief function *fs* is defined via the Shackle identity in the normal way. Now we can redefine *doubt that* as follows:

a doubts that A = {w ∈ W| $fs^{a,w}$(~A) is high}

Contrary to the original interpretation rule, this one does not predict that from *a doubts that A* we can infer that a believes that ~A. Of course we might have done the same by assuming a new accessibility relation, and arguably, that is just what we did. But maybe the accessibility relation ST(a,w) we used has a more general usage than just to help analysing *doubt that*.

6.5 Change by permission

We have discussed earlier in this chapter how desire attributions should be analysed. Now I want to discuss the closely related topic of the dynamics of allowable actions.[366]

Standard deontic logic (SDL) was based on the same principles as classical modal logic and Hintikka's (1962) logic of knowledge and belief. Where normal modal logic has the operators □ and ◊ standing for respectively necessity and possibility, SDL has the two operators O and P. They stand respectively for *ought* or *obliged* and for *permission*. Just like # and ◊ are duals of each other, also P is defined in terms of O as follows: P(A) ≡ ~O(~A) for any proposition A. But this logic has problems to account for *free choice permissions*. Consider the following sentences:

(13) You may take an apple,
(14) You may take a pear, and
(15) You may take an apple or take a pear.

According to SDL, (15) follows from both (13) and (14), and neither (13) nor (14) follows from (15). However, normally if we make a disjunctive permission, both disjuncts are also permitted. The most straightforward way to account for this would be to build it into the formal system by adding an axiom which has that result. Unfortunately, if we assume that P(A ∨ B) ⇒ P(A) ∧ P(B) is a theorem of the logic, we can derive in SDL that everything is allowed. Let us abbreviate P(A ∨ B) ⇒ P(A) ∧ P(B) by FCP. Because in normal deontic logic we can derive O(A ∨ B) from O(A), and P(A ∨ B) follows by definition of P from O(A ∨ B), the assumption that FCP is a theorem of deontic logic has the consequence that from O(A) we can derive P(B) for any A and B; everything is permitted. But, as noted by Kamp (1973), although normally we are allowed to infer from P(A ∨ B) to P(A), and although normally indefinites used in a permission get the arbitrary interpretation, we should not build FCP into our logic because it has counterexamples. Just as for the case of desire, the arbitrarily interpretation of indefinites and disjunctions can be cancelled:

(16) You may pillage city *X* or city *Y*, but first take counsel with my secretary

[366] For another closely related topic, see the various analyses of conditional obligations by Hansson (1969), Van Fraassen (1972), Lewis (1973) and Spohn (1975).

(said by a king to his vassal)
(17) You may pick a flower, but don't pick a rose.

If we cannot stipulate that FCP should be valid, how then can we account for the intuition that normally we conclude from a permission of a disjunction to the permission of its disjuncts?

What we should do, of course, is to follow the lead of the analysis of desire attributions: base the semantic analysis of obligation and permission on a preference relation based on an independent set of propositions. For the analysis of desire, this preference relation was based on the set of propositions that the agent would like to be true. This times the preference relation for the slave should be based on the set of commands given by his master. If John is the slave, let us say that the set of propositions representing the commands given by the master is represented by Per(j,w). I think it is reasonable to assume that this set is mutually consistent, and thus that $\cap$Per(j,w) is non empty. Let us call this set P(j,w). I will assume that this set is a subset of K(j,w), the set of futures consistent with what John believes. It is also natural to assume that the master will only give commands and permissions that are consistent with what the slave believes about the future. On the basis of these assumptions and neglecting anaphora, the interpretation rules for command and permission sentences can be stated in a very simple way:

$[[\text{Must}(j, A)]](S) = \{w \in S \mid P(j,w) \subseteq [[A]](K(j,w))\}$
$[[\text{May}(j, A)]](S) = \{w \in S \mid P(j,w) \cap [[A]](K(j,w)) \neq \emptyset\}$

On the above interpretation rules it is predicted that commands and permissions are closed under logical implication, and that permission sentences don't obey disjunction elimination. And this is the way things should be. As Kamp (1979) observed, there is nothing wrong with the assertion of the following permission sentence:

(18) You may take an apple or take a pear; but I don't know which.

On the other hand, under certain circumstances disjunction elimination is allowed. Let's be a bit more explicit about this. The set P(j,w) gives rise to a reprehensibility relation, $\leq_{P(j,w)}$, for John in w. This time I want to define this ordering relation by following Harper (1976). Thus, $u \leq_{P(j,w)} v$ iff $|S^{w'}{}_u P(j,w)| \geq |S^{w'}{}_v P(j,w)|$, where S is a set of propositions that potentially determine reprehensibility, and w' an arbitrary element of P(j,w).[367] In terms of this preference relation between worlds, we can define a reprehensibility relation between propositions in the following way:

$A \leq_{P(j,w)} B$ iff $\exists u \in A: \forall v \in B: u \leq_{P(j,w)} v$.

Now we can say that from a disjunctive permission of the form Perm(j, $A_1 \vee \ldots \vee A_n$) we are allowed to conclude that the master allows John to do A_i, where $1 \leq i \leq n$, iff for all A_j such that $1 \leq j \leq n$ and $j \neq i$: $A_i \leq_{P(j,w)} A_j$.

It is good to know *when* we can infer from *John may do A or B* to *John may do A and John may do B*. But we want to know something more, we want to know also *why* normally disjunctive permission sentences allow this inference. But this, I think, can not at all be explained by the essentially static way we looked until now at permission sentences. It has been proposed that we can explain this behaviour if we look at the reason why permission sentences are made.

---

[367] Defined in this way, $\leq_{P(j,w)}$ determines a ordering on the worlds that is connected.

There is another phenomenon that we cannot explain until now. Just like desire sentences obey existential weakening, this also holds for permission sentences. We are allowed to infer (20) from (19):

(19) You may eat three apples.
(20) You may eat an apple.

Indeed, this is predicted by the above given account, because permissions are predicted to be closed under logical implication. But what the static account cannot predict is why quantifiers under the scope of *may* get the *at most* reading. Intuitively, after (19) is said by the master John is allowed to eat none, one, two, or three apples, but not more. And after the master said (20), John is still not allowed to eat more than one apple. On the static account, however, this cannot be explained, because quantifiers normally get the *at least* interpretation.

Until now I have looked at commands and permissions from a static perspective. But as noted by Lewis (1970/9) and Kamp (1973, 1979), command and permission sentences are not primarily said to make true assertions, they are rather used by the master to expand or contract the set of the permissible futures for the slave. Kamp (1973) argued that this effect should not be explained in terms of the assertoric use of permission sentences, but in Kamp (1979) he argues with Lewis (1979) and Stalnaker (ms.) that this is probably the best way to think of things. I agree. According to this Lewis/Kamp/Stalnaker account, if the master commands John in *w* to do A by saying *You must do A*, or allows John to do A in *w* by saying *You may do A* it is typically not yet the case in *w* that the proposition expressed by A is respectively a superset of, or consistent with P(j,w). Why else making the command or permission? But in that case both sentences will be false in *w*. But wait! As we learned in chapter 2, we should be realistic about worlds. If a speaker makes an assertion, the world changes. It does not only change because in the new world a certain sentence is uttered by the speaker, but also because in this new world the command or permission sentence will be true. The reason that the command or permission sentence will be true is that the sentence is used performatively, and performatively used sentences can almost never be false (see Lewis, 1970b). But then, for the command or permission sentence to be true, it has to be the case that in the new world, *w'*, the proposition expressed by A is respectively a superset of, or consistent with P(j,w'). Thus, the set representing what is permissible should *indirectly* change from *w* to *w'*, and this indirect change is called *accommodation* by Lewis (1979b). So, the question is what governs the change from P(j,w) to P(j,w').

The most obvious way to go would be to assume that it are both P(j,w) and A that govern this change. And indeed, for commands this seems to be just fine. If the command *You must do A* is given by the master, the set of permissible futures for John in *w'* is intuitively simply $P(j,w) \cap A$. However, things are more complicated for permission sentences. It is clear that if A is allowed, P(j,w') should be a superset of P(j,w) such that $P(j,w') \cap A \neq \emptyset$. It is not clear, however, which A-worlds should be added to P(j,w). Obviously, we cannot simply say that $P(j,w') = P(j,w) \cup A$. By that suggestion, an allowance for A would allow everything compatible with A, which is certainly not what we want. But how then should the change from P(j,w) to P(j,w') be determined if a permission is made? This is Lewis's problem about permissions.

Stalnaker (ms.) suggested that Lewis's problem about permissions can be solved if we change from P(j,w) to P(j,w') due to a permission sentence by *contraction*, where contraction is governed by a reprehensibility ordering on K(j,w). With Harper (1976) we might say that this ordering is determined by the propositions entailed by P(j,w). As I did above, the reprehensibility ordering between worlds is defined as follows, $u \leq_{P(j,w)} v$ iff

$|S^x{}_u P(j,w)| \geq |S^x{}_v P(j,w)|$, where S is a set of propositions that potentially determine reprehensibility, and x an arbitrary element of P(j,w). The change induced by the permission *You may do A* is that in the new world w', $P(j,w') = P(j,w) \cup \{v \in A | \sim\exists u \in A: u <_{P(j,w)} v\}$. But that just means that P(j,w') is the same as $P(j,w) \cup C_{P(j,w)}(A)$.[368] Thus, according to this proposal, command and permission sentences change a context of interpretation as follows:

$[[\text{Must}(j, A)]](S) = \{w' \in W | \exists w \in S: P(j,w') = P(j,w) \cap [[A]] \ \& \ w' \in SSM(w)\}$

$[[\text{May}(j, A)]](S) = \{w' | \exists w \in S: P(j,w') = P(j,w) \cup C''_{P(j,w)}(A) \ \& \ w' \in SSM(w)\}$[369]

Note first that if change by permission is governed by the reprehensibility ordering, the *at most* interpretation of the quantifiers is immediately explained. If John was in w not even allowed to take one single apple, worlds where he takes only apple are closer to the worlds in P(j,w), than worlds where he takes more. So, after the permission that he may take an apple, the new permission set, P(j,w'), can be expected to contain only worlds where he takes *at most* one apple.[370]

Note that if the reprehensibility ordering is governed by Harper's principles, Lewis's objections to this kind of account do not arise. Lewis (1979b) complained that an account in terms of reprehensibility-gradings of worlds might handle single cases of permissions, but leaves undetermined (i) how the comparative permissibility relation evolves from permission to permission, and (ii) how permissibility determines comparative near-permissibility at any given time. But as we have seen, in Harper's construction the permissibility set *does* determine the reprehensibility relation, and thus the change function.

Note that according to this account it does not follow that for a permission sentence of the form *You may do A or B* John can infer that in w' he is allowed to do any of the disjuncts, nor is the arbitrarily interpretation of indefinites guaranteed. To make this a bit more formal, we can say that with respect to P(j,w), A is as least as reprehensible as B, $A \leq_{P(j,w)} B$ iff $\exists u \in \sim A$ and $\forall v \in \sim B$: $|S^x{}_u P(j,w)| \leq |S^x{}_v P(j,w)|$ for any x in K, and where S again is a subset of $\wp(W)$. Then we can say that with respect to P(j,w), A and B are equally strong reprehensible iff $A \leq_{P(j,w)} B$ and $B \leq_{P(j,w)} A$. Because revision functions like our C'' that obey the Gärdenfors postulates satisfy the following factoring condition: $C_K(A \vee B) = C_K(A)$ if $A <_K B$, $C_K(A \vee B) = C_K(B)$ if $B <_K A$, and $C_K(A \vee B) = C_K(A) \cup C_K(B)$ if $A \approx_K B$, we can now explain that normally disjunction elimination is allowed for permission sentences. For simple disjunctive permission sentences like *You may do A or B*, it is not unreasonable to assume, I think, that by a Gricean reasoning we can conclude that the master has no strict preference for the one above the other.

This doesn't mean that this reasoning can be accounted for in a straightforward way. It would be nice to explain the strong reading completely in terms of *conversational implicatures*. Kamp (1979) shows, however, that by the way conversational implicatures are normally understood, as inferences that take as one of their arguments the proposition expressed by sentences, these implicatures can be of no help to explain the strong reading

---

[368] Harper (1977) was the first to define the contraction of K by ~A as $K \cup C''_K(A)$.

[369] This interpretation rule is based on the assumption that there is a set S of propositions that potentially determine reprehensibility. However, things are slightly more complicated; it might be unclear to the slave what this set S is. I will ignore this.

[370] Rohrbauch (1996) claimed that the possible world framework predicts an *at least* reading for quantified permission sentences like (19) and (20). That is not true. Still, I agree with him that what we want for the analysis of permission sentences is that whenever B is a 'natural part' of A, a permission to A 'includes' a permission to B. I later propose an analysis of the notion of 'natural part' that is relevant for the analysis of permission sentences.

of disjunctive permissions. The problem is that these strong readings should also be predicted in case disjunctive permissions are embedded in larger sentences such that the proposition expressed by this larger sentence does not entail the proposition expressed by the embedded disjunctive permission sentence. The following example is given:

(21) Usually you may only take an apple. So if you may take an apple or take a pear, you should bloody well be pleased.

To account for the strong reading of the disjunctive permission it seems that we have to build the implicature into the meaning of *or*, that is that the relevant implicature is not a conversational one, but a *conventional* one instead. But then, how should we account for this conventional implicature? I want to propose that disjunctions of the form $P \vee Q$ can only be appropriately interpreted in contexts K such that $P \cap \sim Q \cap C''_K(P \vee Q) \neq \emptyset$ and $\sim P \cap Q \cap C''_K(P \vee Q) \neq \emptyset$. As a result, it is predicted that a normal sentence of the form *A or B* can only be appropriately asserted in a context that is compatible with both $A \wedge \sim B$ and $\sim A \wedge B$, and that a permission sentence like *You may do A or B* can only be said appropriately by the master to John in *w* iff $A \approx_{P(j,w)} B$.

In the discussion of Kamp (1979) it seemed that only disjunctive permission sentences give us trouble. But as Merin (1992) noticed, also *conjunctive* permission sentences are problematic. In the original Stalnaker account it is predicted that a conjunctive permission sentence *You may take an apple and a pear* has semantically the package deal effect: take either none or both. The reason is that $C''_{P(j,w)}(A \wedge B)$ is a subset of $A \cap B$. But this package deal effect is empirically wrong, since a conjunctive permission allows also for the conjuncts to be done separately.

How can we get rid of the package deal effect for conjunctive permissions, and still interpret $\wedge$ as intersection? On first thought there seems to be an easy solution, analyse permission sentences in the following way:

$$[[\text{May}(j, A)]](S) = \{w' \mid \exists w \in S: P(j,w') = \cup\{C''_{P(j,w)}(B) \mid A \subseteq B\} \;\&\; w' \in SSM(w)\}$$

Note that the weaker the proposition is, the closer the selected world(s) will be to the world where we start from, thus $A \subseteq B \Rightarrow B \leq_{P(j,w)} A$. Because both A and B follow from $A \wedge B$, it is predicted that if it is allowed that $A \wedge B$, also A and B alone are allowed. However, it should be clear that this can't be good enough. It is wrongly predicted that everything that is less reprehensible than A is allowed once A is allowed. The reason is simple, because for any A and B, $A \subseteq A \vee B$, it follows that $C''_{P(j,w)}(A \vee B) \supseteq C''_{P(j,w)}(B)$ if $B \leq_{P(j,w)} A$, and thus also the new permission set after the permission of A will contain B-worlds and is thus allowed.

The problem we face seems to be a familiar one: sometimes conjunction seems to behave like a disjunction. But this problem has an equally familiar solution: *lifting*. That is, we can treat $\wedge$ still as intersection, but then not on the normal meaning of its two conjuncts, but on their lifted reading. Let F_A be the principle filter generated by A: $\{C \subseteq W \mid A \subseteq C\}$, then it will be the case that $F_A \cap F_B = F_{A \vee B}$. Then we can say that for the permission sentences, we should not look primarily at the normal semantic value, but at the filter generated by its normal semantic value, instead. But not just that. We also have to say that the semantic value of the embedded clauses of permission sentences are always different from their normal semantic value, and even in such a way that this other semantic value is

not derivable from it's normal semantic value. I'm not saying that this kind of move cannot be made, but it is not straightforward, and in need of explanation.

A completely different solution to the problem would be to suggest that the embedded clauses of permission sentences do not express propositions, but *actions* modelled by *dynamic programs* instead. For simplicity we might think of programs as functions from worlds to sets of sets of worlds, and if *a* is an atomic or negated atomic program, $|a|_w$ denotes the following singleton set: $\{f_w(A)\}$, where $f_w(A)$ is the image of *w* under *A*, the set of closets worlds to *w* where A is the case. The assumption can be made that the new permission state after the permission of A, P(a,w') is simply $P(a,w) \cup \cup\{|a|_{w'} \mid w' \in P(a,w)\}$. However, this won't quite do when we also consider disjunctive and *conjunctive programs*. I will assume that when *A* and *B* denote programs, *A and B* either denotes *a;b* the execution of *b* after *a*, or the conjunctive program $a \wedge b$. Disjunctive and conjunctive programs denote the following sets of sets:

$|a \vee b|_w = \{\cup|a|_w, \cup|a|_w\}$

$|a ; b|_w = \{\cup|b|_{w'} \mid w' \in \cup\cup|a|_w\}$

$|a \wedge b|_w = |a|_w \cdot |b|_w$,

where $X \cdot Y = \{e \cdot e' \mid e \in X \,\&\, e' \in Y\}$, and if e and e' are sets of possible worlds, then $e \cdot e' = e \cup e'$.

Thus, if π denotes a complex program, $|\pi|_w$ might be a set that contains more than one set of possible worlds. For instance, if *A*, *B* and *C* are atomic program denoting sentences, and if $f_w(A) = \{u,u',u''\}$, $f_w(B) = \{v,v'\}$, and $f_w(C) = \{w',w''\}$, then $|a \vee b|_w = \{\{u,u',u''\}, \{v,v'\}\}$ and $|(a \vee b) \wedge c|_w = \{\{u,u',u'',w',w''\}, \{v,v',w',w''\}\}$.

Now we can interpret permission sentences as follows:

$[[\text{May}(j, A)]](S) = \{w' \in W \mid \exists w \in S: w' \in SSM(w) \,\&\, P(j,w') = P(j,w) \cup \cup\{B \in |a|_w \mid \sim\exists C \in |a|_w: C <_{P(j,w)} B\}$.

Thus, the permission set of John in *w* changes in such a way that the best elements of $|a|_w$ with respect to the reprehensibility ordering induced by the old permission set, P(j,w), are allowed. Because it will be the case that if *A* and *B* are atomic program denoting sentences, $|a \wedge b|_w$ will be a singleton set, but $|a \vee b|_w$ not, it is predicted that such conjunctive permission sentences always allow for both (probably done separately), but that this doesn't hold for disjunctive permission sentences

Although the above solution seems to predict quite good, you might prefer a solution to our problem that gives a semantic value to the embedded clause A from which we can derive its normal semantic value, and the value needed to account for permission sentences. Instead of using conjunctive programs to solve the problem, I will assume the existence of *conjunctive facts*.[371] I will argue later that there might be such a (Russellian) semantic value, and that by means of this value we can define a strange kind of consequent relation, $\models_{EM}$, such that we can interpret permission sentences as follows:

---

[371] Actually, the proposal using conjunctive programs was partly motivated by Van Fraassen's (1969) account of conjunctive facts to be discussed below.

$$[[\text{May}(j, A)]](S) = \{w' \in W \mid \exists w \in S: P(j,w') = P(j,w) \cup \cup\{C''P(j,w)(B) \mid A \models_{EM} B\} \ \& \ w' \in SSM(w)\}$$

Fortunately, the consequence relation will be reflexive, and conjunctive elimination is allowed. But $\models_{EM}$ has a strange property. The consequence relation will be very deviant because $A \vee B$ will not follow from A if B does not denote a subfact of A, or the other way round.

How strange this consequence relation might be, if the relation has these formal properties, then, I think, the predictions made by the above interpretation rule are satisfying. First, for disjunctions basically the same predictions are made as earlier, and second, for a conjunctive permission of the form *You may do A and B*, not only $A \wedge B$, but also A separate and B separate are allowed.

Until now it is not at all clear how we could think of a consequence relation that has these properties. One of the goals of the next section will be to define this consequence relation.

6.6 Facts and factives

In the first chapter I followed Stalnaker in arguing that the content of belief should be accounted for in externalistic terms. The content of one's belief should be defined in terms of nested counterfactuals, being dependent on normal or optimal conditions. Although the content depends crucially on external factors, it is not the actual causal relation between object and representation that primarily counts for the analysis of belief. For the analysis of factive verbs like *know* and *regret* things seem to be different. For the analysis of these verbs we should not look at the more general counterfactual relations between the representational system of the agent and his environment, but at the more specific causal relations that exists between the two. This, I think, is what Gettier's problem for the traditional analysis of knowledge as justified true belief,[372] and Vendlerian arguments suggesting that more concrete entities than propositions are relevant for the analysis of knowledge.[373] Indeed, Ginzburg (1994) proposed on the basis of Vendlerian arguments that situation semantics is a better framework for the analysis of knowledge than the possible world framework that I used in this dissertation, and Kratzer (ms) argued that the Gettier problem can be solved if knowledge is analysed as justified *de re* belief about facts. I agree with Kratzer, and now want to suggest how we can account for Kratzer's suggestion in the framework of the double indexing counterpart modal logic stated in chapter 1. For this I will use one of the earliest formal analyses of facts, a system proposed by Van Fraassen in 1969. In terms of the apparatus developed by him we will also define the consequence relation that we used for the analysis of permission sentences above.

Van Fraassen bases his analysis of facts on the Russellian notions of complexes and facts. A complex is any $n + 1$ tuple whose first member is an n-ary relation on D and whose other members are members of D. A fact in a model is any non-empty set of complexes in the model. A complex is something that makes an atomic sentence true. I assume that for every complex $\langle \mathbf{R}, d_1, \ldots, d_n \rangle$ that verifies a sentence of the form $R(a_1,\ldots,a_n)$ in a world, there is also a complex $\langle \mathbf{cR}, d_1, \ldots, d_n \rangle$ that makes $R(a_1,\ldots,a_n)$ false in a world. So, the fact that corresponds with the sentence $R(a_1,\ldots,a_n)$ is $\{\langle \mathbf{R}, d_1, \ldots, d_n \rangle\}$. On the Russellian assumption that a sentence *A* is true (false) iff some fact that makes *A* true (false) is the case, Van Fraassen assumes that except atomic facts that make atomic sentences true, we also need atomic facts that make atomic sentences false, and conjunctive facts that make conjunctions true. We designate the union of facts $f_1,\ldots,f_n$, by $f_1 \cdots f_n$ and this a *conjunctive fact* with components $f_1,\ldots,f_n$. We say that f *forces* f' if both f and f' are facts,

[372] Unsurprisingly, the arguments very much resemble the arguments against the description theory of meaning.
[373] See also Asher (1993).

and *f'* is a subset of *f*. If we want to use the model theory of chapter 1, we say that models are just as they are there except that for every world *w*, we also assume there to be primitive relations $\mathbf{R_1}$, $\mathbf{R_2}$,... on the domain of *w*. I assume that for every pair of worlds *w* and *w'*, and for every primitive predicate R_i in the language there are two relations $\mathbf{R_i}$ and $\mathbf{cR_i}$, such that $I^+_{w,w'}(R_i) = \mathbf{R_i}$, and $I^-_{w,w'}(R_i) = \mathbf{cR_i}$.

If *X* and *Y* are two sets of facts, we will say that $X \cdot Y$ is the *product* of *X* and *Y* determined as follows: $X \cdot Y = \{f \cdot f' | f \in X \text{ and } f' \in Y\}$.

$T_{w,w',c,g}(A)$ will be the set of minimal facts in w' that make A true in w' with respect to w, c and g. A similar thing holds for $F_{w,w',c,g}(A)$.

$T_{w,w',c,g}(Rt_1,...,t_n)$ = $\{\{<\mathbf{R}, \|t_1\|^{w,w',c,g},..., \|t_n\|^{w,w',c,g})>\} | <\|x_1\|^{w,w',c,g},..., \|x_n\|^{w,w',c,g})> \in I^+_{w,w',c,g}(R)\}$, if P is an atomic predicate

$F_{w,w',c,g}(Rt_1,...,t_n)$ = $\{\{<\mathbf{cR}, \|t_1\|^{w,w',c,g},..., \|t_n\|^{w,w',c,g})>\} | <\|x_1\|^{w,w',c,g},..., \|x_n\|^{w,w',c,g})> \in I^-_{w,w',c,g}(R)\}$, if P is an atomic predicate

$T_{w,w',c,g}(\sim A) = F_{w,w',c,g}(A)$

$F_{w,w',c,g}(\sim A) = T_{w,w',c,g}(A)$

$T_{w,w',c,g}(A \wedge B) = T_{w,w',c,g}(A) \cdot T_{w,w',c,g}(B)$

$F_{w,w',c,g}(A \wedge B) = F_{w,w',c,g}(A) \cup F_{w,w',c,g}(B)$

$T_{w,w',c,g}(\forall x A)$ = the product of the sets $T_{w,w',c,g[x/d]}(A)$ for any $d \in D(w)$

$F_{w,w',c,g}(\forall x A) = \cup\{F_{w,w',c,g[x/d]}(A) | d \in D(w)\}$

The formulae '$A \vee B$', and '$\exists x A$' will be treated as abbreviations for '$\sim(\sim A \wedge \sim B)$' and '$\sim\forall x \sim A$', respectivily.

For illustration, suppose that $T_{w,w',c,g}(A) = \{\{<\mathbf{R},a,b>\}\}$, $T_{w,w',c,g}(B) = \{\{<\mathbf{P},c>\}\}$, and $T_{w,w',c,g}(C) = \{\{<\mathbf{Q},d>\}, \{<\mathbf{S},e>\}\}$. Then $T_{w,w',c,g}(A \wedge B \wedge C) = \{\{<\mathbf{R},a,b>,<\mathbf{P},c>, <\mathbf{Q},d>\}, \{<\mathbf{R},a,b>,<\mathbf{P},c>, <\mathbf{S},e>\}\}$, and $T_{w,w',c,g}(A \vee B \vee C) = \{\{<\mathbf{R},a,b>\},\{<\mathbf{P},c>\}, \{<\mathbf{Q},d>\}, \{<\mathbf{S},e>\}\}$.

Whatever facts are, a world is a maximally consistent conjunctive fact.[374] A fact obtains in a world, if the world forces this fact.

Obt(f, w) iff Force(w,f) iff $f \subseteq w$

A sentence is true (false) in a world with respect to a context world, a counterpart function and an assignment iff there is a fact that obtains in this world that is among the truth (falsity) set of A with respect to this context world, a counterpart function and an assignment.

$w,w',c,g \models A$ iff $\exists f \in T_{w,w',c,g}(A)$: Obt(f,w')

$w,w',c,g \dashv A$ iff $\exists f \in F_{w,w',c,g}(A)$: Obt(f,w')

---

[374] To define worlds in this way, we need to know what maximality is and a primitive notion of consistency.

The classical entailment relation is defined as expected (where I assume that ç is the only metaphysical counterpartfunction):

$A \models_{g,s,w} B \quad \text{iff} \quad \{w' \in W | w,w',ç,g \models A)\} \subseteq \{w' \in W | w,w',ç,g \models B\}$

But we did not introduce facts to account for truth and classical entailment, we don't need facts to account for that. We introduced them in order to analyse factive verbs, and for a semantic analysis of our strange entailment relation $\models_{EM}$. But let us first discuss a more familiar kind of entailment relation. To be able to do that I define the notions $T^*_{w,w',c,g}(A)$ and $F^*_{w,w',c,g}(A)$, being the set of all facts that force some fact in $T_{w,w',c,g}(A)$ and $F_{w,w',c,g}(A)$, respectively:

$T^*_{w,w',c,g}(A) = \{f \subseteq w' | \exists f' \in T_{w,w',c,g}(A): f' \subseteq f\}$
$F^*_{w,w',c,g}(A) = \{f \subseteq w' | \exists f' \in F_{w,w',c,g}(A): f' \subseteq f\}$[375]

Now we can define the following notion of entailment:

$A \models_{g,w,t} B \quad \text{iff} \quad \forall w' \in W: T^*_{w,w',ç,g}(A) \subseteq T^*_{w,w',ç,g}(B)$
$\text{iff} \quad \forall w' \in W: \forall f \in T_{w,w',ç,g}(A): \exists f' \in T_{w,w',ç,g}(B): f' \subseteq f$

Van Fraassen (1969) proves that for the propositional system it holds that $A \models_{g,w,t} B$ iff B is a *tautological entailment* of A.[376]

Another, more deviant, consequence relation can be defined as follows:

$A_{g,w} \models_{EM} B \quad \text{iff} \quad \forall w' \in W: \forall f \in T_{w,w',ç,g}(A): \exists f' \in T_{w,w',ç,g}(B): f' \subseteq f \quad \text{and}$
$\forall w' \in W: \forall f' \in T_{w,w',ç,g}(B): \exists f \in T_{w,w',ç,g}(A): f' \subseteq f$

Intuitively, only parts of A stand in this consequence relation with A. This is the notion that I needed for the analysis of permission sentences.[377]

To define this consequence relation was one of the goals of this section, to be able to give a hint at how we could give a *de re* analysis of a factive verb like *knowledge* is another. In order to do that we have to assume that facts that obtain in a world are special kinds of objects of the domain of that world that variables can range over. Furthermore I assume that for every formula *A*, 'A' is a singular term, and that for every *w*, *w'*, *c* and *g*, $[\![\text{'A'}]\!]^{w,w',c,g} = A$, and $I^+_{w,w',c,g}(\text{Fact}) = \{\langle A,f\rangle | A \in \text{FORM}_L \ \& \ f \in T_{w,w',ç,g}(A)\}$ and $I^-_{w,w',c,g}(\text{Fact}) = \{\langle A,f\rangle | A \in \text{FORM}_L \ \& \ f \in F_{w,w',ç,g}(A)\}$.[378] We could now interpret knowledge as a *de re* attitude in the following way:

John knows that A $\quad \rightarrow \quad \exists \hat{x}\,(\text{Bel}(j, ¡\text{Fact}('A',x) \wedge \text{Obt}(x)))$

---

[375] Note that it also holds that

$w,w',c,g \models A \quad \text{iff} \quad \exists f \in T^*_{w,w',c,g}(A): \text{Obt}(f,w')$
$w,w',c,g \dashv A \quad \text{iff} \quad \exists f \in F^*_{w,w',c,g}(A): \text{Obt}(f,w')$

[376] The notion of *tautological entailment* is due to Anderson & Belnap (1962). When we restrict ourselves to only one possible world, the relation comes down to the notion called *lumping* in Kratzer (1989).

[377] However strange the consequence relation might be, it turns out that the relation already has a name. The ordering between sets triggered by $\models_{EM}$ is known as the *Egli-Milner* preorder in Computer Science (Tim Fernando, personal communication). Note that it doesn't hold that $P(a)_{g,w} \models_{EM} \exists x Px$, an inference Soames (1987) wants to block for the analysis of belief attributions by means of Russellian propositions.

[378] This all makes the notion of a fact much too language dependent, but I won't bother here.

Note that for this *de re* analysis I make use of the double indexing account also used in the more traditional analysis of embedded questions. The above formula can only be interpreted if we know what the relevant predicates mean:

$T_{w,w',c,g}(Obt(t)) = \{v \in W | v = w' \ \& \ [[t]]^{w,w',c,g} \subseteq w'\}$
$F_{w,w',c,g}(Obt(t)) = \{v \in W | v = w' \ \& \ [[t]]^{w,w',c,g} \not\subseteq w'\}$

$T_{w,w',c,g}(Fact('A',t)) = \{v \in W | v = w' \ \& \ [[t]]^{w,w',c,g} \in T_{w,w',c,g}(A)\}$
$F_{w,w',c,g}(Fact('A',t)) = \{v \in W | v = w' \ \& \ [[t]]^{w,w',c,g} \notin T_{w,w',c,g}(A)\}$

If P is a complex predicate of the form $\hat{x}$ A (where A $\in$ $FORM_L$),
then (1) $I^+_{w,w',c,g}(P) = \{d \in D(w') : T_{w,w'',c',g[x/d]}(A) \neq \emptyset\}$;
(2) $I^-_{w,w',c,g}(P) = \{d \in D(w') : F_{w,w'',c',g[x/d]}(A) \neq \emptyset\}$.

For the analysis of belief, and thus for knowledge in terms of it, I will assume that belief attributions are formalised in the following form: Bel(t, 'A'). The set $T_{w,w',c,g}$(Bel(t, 'A')) is not empty iff a fact of the form {<**BEL**, $[[t]]^{w,w',c,g}$, A>} obtains in w'. Such a fact obtains in *w'* iff $<[[t]]^{w,v',c,g}$, A> $\in$ $I^+_{w,w',c,g}$(Bel), where $I^+_{w,w',c,g}(Bel) = \{<d,A> | \exists c' \in C_{acq}(d,w') : \forall w'' \in K(d,w') : T_{w,w'',c,g}(A) \neq \emptyset\}$.[379] Now it follows per definition that a belief sentence is true iff there is a fact that makes it true. Something similar can be done for falsity of belief sentences.

Note that in the way we have analysed knowledge attributions, a disjunctive knowledge attribution like *Smith knows that either Jones owns a Ford, or Brown is in Barcelona* (Gettier, 1963) will only be counted as true if either Jones owns a Ford, and Smith believes this in a *de re* way, or Brown is in Barcelona, and Smith believes this by knowing a particular fact that makes this true. Moreover, if we would analyse *John knows whether A* as *John knows that A or not-A* we would predict that if A is true, John must have a *de re* belief about an actual fact that makes A true, and if ~A is true, John must have a *de re* belief about an actual fact that makes ~A true.

In this dissertation I tried very hard to account for phenomena in terms of possible worlds only. Now I suggested that we need facts or situations, too. It's high time to stop.

[379] This, of course, also gives rise to foundational problems which I don't want to get into here.

Literature

Abusch, D., (ms.), *Sequence of Tense Revisited*, University of Stuttgart.

Adams, E.W. (1965), "On the logic of conditionals", *Inquiry*, 8, pp.166-197.

Adams, E.W. (1970), "Subjunctive and indicative conditionals", *Inquiry*, 6, pp. 39-94.

Adams, E.W. (1976), "Prior probabilities and counterfactual conditionals", In: W.L. Harper and C.A. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol I, Reidel, Dordrecht.

Alchourrón, C.E. *et al.* (1985), "On the logic of theory chnage: Partial meet functions for contraction and revision.", *Journal of Symabolic Logic*, 50, pp. 510-530.

Anderson, A.R. and N.D. Belnap (1962), "Tautological entailments", *Philosophical Studies*, 13, pp. 19-52.

Asher, N. (1986), "Belief in discourse representation theory", *Journal of Philosophical Logic*, 15, pp. 127-189.

Asher, N. (1987), "A typology for attitude verbs and their anaphoric propertics", *Linguistics and Philosophy*, 10, pp. 125-197.

Asher, N. (1993), *Reference to Abstract Objects in Discourse*, Dordrecht, Kluwer Academic Publishers.

Bäuerle R. and M. J. Cresswell, (1984), "Propositional attitudes", In: D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic, Vol. II*, D. Reidel, Dordrecht.

Beaver, D.I. (1992), "The kinematics of presupposition", In: H. Kamp (ed.), *Presupposition*, ILLC University of Amsterdam, Amsterdam, Dyana-2 deliverable R2.2.A, part II.

Beaver, D.I. (1993), *What comes first in dynamic semantics*, ILLC report LP-93-15, University of Amsterdam.

Beaver, D.I. (1995), "Accommodating topics", In: H. Kamp and B. Partee (eds.), *Context in the Analysis of Linguistic Meaning*, Stuttgart/Prague.

Beaver, D.I. (1996), "Presupposition", In: J. van Benthem and A. ter Meulen, (eds.), *Handbook of Logic and Language*, Elsevier.

Belnap, N.D. (1970), "Conditional assertion and restricted quantification", *Nous*, 4, pp. 1-12.

Blok, P.I. (1993), *The interpretation of Focus, An epistemic approach to pragmatics*, Ph.D. dissertation, University of Groningen.

Bochvar, D.A. (1939), "Ob odsom trehznachom iscislenii i ego primeneii k analizu paradoksov klassicskogo rassirennoga funkcional 'nogo iscislenija'" *Mathematiciskij sbornik*, 4.

Bratman, M.E. (1987), *Intention, Plans, and Practical Reason*, Cambridge: Harvard University Press.

Burge, T. (1979), "Individualism and the mental", In: P. Frech *et al.* (eds.), *Midwest Studies in Philosophy, 4, Studies in Epistemology*, Mineapolis, University of Minnesota Press.

Buridan, J. (1350), *Sophismata*, trans. T.K. Scott as *Sophisms on Meaning and Truth*, Appleton-Century-Crofts, 1966.

Carnap, R. (1947), *Meaning and Necessity. A study in Semantics and Modal Logic*, Chicago.

Chastain, C. (1975), "Reference and context", In: K. Gunderson (ed.), *Menesota Sudies in the Philosophy of Science, vol. VII- Language, Mind, and Knowledge*, Minneapolis: University of Menneasota Press.

Chierchia, G. (1989), "Anaphora and Attitudes de se", In: R. Bartsch *et al.* (eds.), *Semantics and Contextual Expressions*, Dordrecht.

Chierchia, G. (1996), *Dynamics of Meaning, Anaphora, Presuppositions and the Theory of Grammar*, University of Chicago Press, Chicago.

Chisholm, R. (1967), "Identity through possible worlds: some questions", *Nous*, 1. pp 1-18.

Church, A. (1943), "Review of Quine", *Journal of Philosophical Logic*.

Church, A. (1954), "Intensional isomorphism and identity of belief", *Philosophical Studies*, 5, ,pp. 65-73.

Church, A. (1982), "A remark concerning Quine's paradox about modality", Spanish version in *Analysis Filosofico*, pp. 25-32.
Cooper, R.H. (1979), "The interpretation of pronouns", In: F. Heny and H.S. Schnelle (eds.), *Selections from the third Groningen round table, Syntax and Semantics*, 10, Academic Press, New York.
Cresswell, M. (1973), *Logics and Languages*, Methuen, London.
Cresswell, M. and A. von Stechow, (1982), "De re belief generalised", *Linguistics and Philosophy*, 5, pp. 503-535.
Crimmins, M and J. Perry, (1989), "The prince and the phone booth: reporting puzzling beliefs", *Journal of Philosophy*, LXXXVI, pp. 685-711.
Deemter, K. van (1991), *On the Composition of Meaning*, Ph.D. dissertation, University of Amsterdam.
Dekker, P. (1993), *Transsentential Meditations, Ups and downs in dynamic semantics*, Ph.D. dissertation, University of Amsterdam.
Dekker, P. (1994), "Predicate logic with anaphora", In: R. Cooper and J. Groenendijk, *Integrating Semantic Theories II, Dyana-2*, Deliverable R2.1.B.
Dekker, P. (1995), "On context and identity", In: H. Kamp and B. Partee (eds.), *Context in the Analysis of Linguistic Meaning*, Stuttgart/Prague, 1995.
Dekker, P. (1996), "Reference and representation", In: K. von Heusinger and U. Egli (eds.), *Proceedings of the Konstanz Workshop "Reference and Anaphoric Relations"*, Konstanz.
Dennett, D.C. (1969), *Content and Consciousnes*, London, Routledge and Kegan Paul.
Does, J. van der. (1994), "Formalising E-type logic", In: P. Dekker and M. Stokhof (eds.), *Proceedings of the Ninth Amsterdam Colloquium*, Amsterdam, pp. 229-248.
Donnellan, K. (1966), "Reference and definite descriptions", *Philosophical Review*, 75, pp. 281-304.
Donnellan, K. (1970), "Proper names and identifying descriptions", *Synthese*, 21, pp. 3- 31.
Donnellan, K. (1974), "Speaking of nothing", *Philosophical Review*, 83, pp. 3-32.
Donnellan, K. (1978), "Speaker reference, descriptions, and anaphora", In: P. Cole (ed.), *Syntax and Semantics, vol. 9: Pragmatics*, New York: Academic Press, pp. 47-68.
Dretske, F.L. (1970), "Epistemic operators", *The Journal of Philosophy*, 67, pp. 1007-1023.
Dretske, F.L. (1981), *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press/Bradford Books.
Edelberg, W. (1986), "A new puzzle about intentional identity", *Journal of Philosophical Logic*, 15, pp. 1-25.
Edelberg, W. (1992), "Intentional identity and the attitudes", *Linguistics and Philosophy*, 15, pp. 561-596.
Edelberg, W. (1995), "A perspectival semantics for the attitudes", *Nous*, 29:3, pp. 316-342.
Eijck, J. van and G. Ceparello, (1994), "Dynamic modal predicate logic", In: M. Kanazawa and C. Pinon (eds.), *Dynamics, Polarity and Quantification*, CSLI, Stanford, pp. 251-276.
Evans, G. (1973), "The causal theory of names", *Proceedings of the Aristotelian Society, Supplementary Volume* 47, pp. 84-104.
Evans, G. (1977), "Pronouns, quantifiers and relative clauses (1)", *The Canadian Journal of Philosophy*, 7, pp. 467-536.
Evans, G. (1979), "Reference and contingency", *The Monist*, lxii, pp. 161-189.
Evans, G. (1981), "Understanding demonstratives", In: Parret and Bouveresse (eds.), *Meaning and Understanding*, Belin, pp. 280-303.
Evans, G. (1982), *Varieties of Reference*, Oxford University Press.
Fagin and Halpern, (1988), "Belief, awareness and limited reasoning", *Artificial Intelligence*, 34, pp. 39-76.
Fauconnier, G. (1984), *Mental Spaces*, Cambridge Mass: MIT Press.

Fernando, T. (1994), "Generalised quantifiers as second-order programs - "dynamically" speaking, naturally", In: P. Dekker and M. Stokhof (eds.), *Proceedings of the Ninth Amsterdam Colloquium*, ILLC, Amsterdam.
Fernando, T. (1995), "Are context change potentials functions?", In: H. Kamp and B. Partee (eds.), *Context in the Analysis of Linguistic Meaning*, Stuttgart/Prague.
Fine, K. (1975), "Review of Lewis's 'Counterfactuals'", *Mind*, 84, pp. 451-458.
Fine, K. (1983), "A defence of arbitrary objects", *Proceedings of the Aristotelian Society*, Supplementary Volume LVII, pp. 5-77.
Fintel, K. von (1994), *Restrictions on Quantifier Domains*, Ph.D. dissertation, University of Massachusetts, Amherst.
Fintel, K. von (1995), "Presupposition Accommodation and Quantifier Domains. Comments on Beaver's 'Accommodating topics'", In: H. Kamp and B. Partee (eds.), *Context in the Analysis of Linguistic Meaning*, Stuttgart/Prague.
Fodor, J.A. (1987), *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press/Bradford Books.
Fodor, J.D. (1979), *The Linguistic Description of Opaque Contexts*, Garland, New York.
Fodor, J.D. and I.V. Sag, (1982), "Referential and quantificational indefinites", *Linguistics of Philosophy*, 5, pp. 355-398.
Forbes, G. (1985), *The Metaphysics of Modality*, Clarendon Press, Oxford.
Fraassen, B.C. van, (1966), "Singular terms, truth-value gaps, and free logic", *Journal of Philosophy*, 63, pp. 481-495.
Fraassen, B.C. van, (1969), "Facts and tautological entailments", *Journal of Philosophy*, 66, pp. 477-487.
Fraassen, B.C. van (1972), "The logic of conditional obligation", *Journal of Philosophical Logic*, 1, pp. 417-438.
Fraassen, B.C. van (1974), "Hidden variables and conditional logic", *Theoria*, 40.
Fraassen, B.C. van (1976), "Probabilities of conditionals", In: W.L. Harper and C.A. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science,* Volume I, Reidel, Dordrecht.
Fraassen, B.C. van, (1977), "The only necessity is verbal necessity", *Journal of Philosophy*, 74, pp. 71-85.
Fraassen, B.C. van, (1979), "Propositional attitudes in weak pragmatics", *Journal of Philosophy*, 76, pp. 365-374.
Fraassen, B.C. van (1980), *The Scientific Image*, Oxford University Press.
Frank, A. (1997), *Context Dependence in Modal Constructions*, Ph.D. dissertation, University of Stuttgart.
Frege, G. (1892), "Über Sinn und Bedeutung", *Zeitschrift für Philosophie und philosophische Kritik*, 50, pp. 25-50.
Gamut, L.T.F. (1991), *Logic, Language and Meaning, Vol 2: Intensional logic and logical grammer*, Chicago: The univeristy of Chicago Press.
Gärdenfors, P. (1982), "Imaging and conditionalisation", *Journal of Philosophical Quarterly*, 18, pp. 203-211.
Gärdenfors, P. (1988), *Knowledge in Flux, Modeling the Dynamics of Epistemic States*, Cambridge Mass., MIT Press.
Gärdenfors, P. and D. Makinson, (1994), "Nonmonotonic inference based on expectations", *Artificial Intelligence*, 65, pp. 197-245.
Garson, J.W. (1984), "Quantification in modal logic", In: D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic, Vol. II*, D. Reidel, Dordrecht, pp. 249-307.
Gazdar, G. (1979), *Pragmatics, Implicatures, Presuppositions, and Logical Form*, Academic Press, New York.
Geach, P. (1962), *Reference and Generality*, Ithaca, NY: Cornell University Press.
Geach, P. (1967), "Intentional identity", *Journal of Philosophy*, 64, pp. 627-632.
Gettier, E. (1963), "Is justified true belief knowledge?", *Analysis*, 6, pp. 121-123.
Geurts, B. (1995), *Presupposing*, Ph.D. Dissertation, University of Stuttgart.
Gibbard, A. (1975), "Contingent identity", *Journal of Philosophical Logic*, 4, pp.187-222.
Gibbard, A. (1980), "Two recent theories of conditionals", In: W. L. Harper *et al.* (eds), *Ifs*, Reidel, Dordrecht.

Gibbard, A. and W.L. Harper (1978), "Counterfactuals and two kinds of expected utility", In: C. Hooker *et al.* (eds.), *Foundations and Applications of Decision Theory*, Western Ontario Series in the Philosophy of Science, Vol. 1, Reidel, Dordrecht.

Gillon, B.S. (1996), "Three theories of anaphora and a puzzle from C.S. Peirce", In P. Dekker and M. Stokhof (eds.), *The Proceedings of the 10'th Amsterdam Colloquium*, ILLC, Univerisity of Amsterdam.

Grice, H.P. (1967), "Logic and conversation", Willian James Lectures, Harvard. Published in D. Davidson and G. Harman (eds.), 1976, *The Logic of Grammar*.

Groenendijk, J. and Stokhof, M. (1982), "Semantic analysis of Wh-complements", *Linguistics and Philosophy*, 5, pp. 175-233.

Groenendijk, J. and M. Stokhof (1984), *Studies on the Semantics of Questions and the Pragmatics of Answers*, Amsterdam.

Groenendijk, J. and M. Stokhof (1990), "Dynamic Montague Grammar", In: L. Kalman and L. Polos (eds.), *Papers from the second symposium on logic and language*, Budapest, 1990. Adademia Kiado.

Groenendijk, J. and M. Stokhof (1991), "Dynamic predicate logic", *Linguistics and Philosophy*, 14, pp. 39-100.

Groenendijk, J. *et al.* (1994), "Dynamic semantics", *Course material for the Sixth European Summerschool on Logic, Language and Information*. Copenhagen.

Groenendijk, J. *et al.* (1995a), "Coreference and modality in the context of multi-speaker discourse", In: H. Kamp and B. Partee (eds.), *Context in the Analysis of Linguistic Meaning*, Stuttgart/Prague.

Groenendijk, J. *et al.* (1995b), "Coreference and contextually restricted quantification. Is there another choice?", In: H. Kamp and B. Partee (eds.), *Context in the Analysis of Linguistic Meaning*, Stuttgart/Prague.

Grove, A.J. (1986), *Two Modellings for Theory Change*, Auckland Philosophy Papers 13.

Grove, A.J. (1995), "Naming and identity in epistemic logic part 2: A first-order logic for naming", *Artificial Intelligence*, vol. 74, nr. 2. pp. 311-350.

Haas-Spohn, U. (1986), "Zur interpretation der Einstellungszuschreibungen", *Sonderforschungsbereich 99*, University of Konstanz.

Haas-Spohn, U. (1994), *Versteckte Indexikalitat und subjective Bedeutung*, Ph.D. Dissertation, University of Tübingen.

Hajek, A. and N. Hall, (1994), "The hypothesis of the conditional construal of conditional probabilities", In: E. Eells and B. Skyrms (eds), *Probability and Conditionals*, Cambridge University Press, Cambridge.

Hamblin, C.L. (1973), "Questions in Montague English", *Foundations of Language*, 10, pp. 41-53.

Hansson, B. (1969), "An analysis of some deontic logics", *Nous*, 3, pp. 373-398.

Hansson, S. O. (1989), "A new semantical approach to the logic of preference", *Erkenntnis*, 31, 1-42.

Harper, W.L. (1975), "Rational belief change, Popper functions, and counterfactuals", *Synthese*, 30, 221.

Harper, W.L. (1976), "Ramsey test conditionals and iterated belief change (A response to Stalnaker)", In: W.L. Harper and C.A. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol I, Reidel, Dordrecht.

Harper, W.L. (1977), "Rational conceptual change", In: *PSA 1976*. East Lansing Mich.: Philosophy of Science Association, vol. 2, 462-135.

Harper, W.L. (1980), "A sketch of some recent developments in the theory of conditionals", In: W.L. Harper *et al.* (eds.), *Ifs*, Reidel, Dordrecht.

Hazen, A. (1979), "Counterpart theoretic semantics for modal logic", *The Journal of Philosophy*, 76, pp. 319-338.

Heim, I. (1979), "Conceiled questions", In: R. Bartsch *et al.* (eds.), *Semantics from different points of view, Language and communication 6*, Springer, Berlin, pp. 61-74.

Heim, I. (1982), *The Semantics of Definite and Indefinite Noun Phrases*, Ph.D. dissertation, University of Massachusets, Amherst.

Heim, I. (1983), "On the projection problem for presuppositions", In: M. Barlow *et al.* (eds.), *West Coast Conference on Formal Linguistics*, pp. 114-123.
Heim, I. (1990), "E-type pronouns and donkey anaphora", *Linguistics and Philosophy*, 13, pp. 137-178.
Heim, I. (1991), "Artikel und definiteit", In: A. von Stechow and D. Wunderlich (eds.), *Semantik: ein internationales Handbuch der zeitgenossischen Forschung*, Berlin, pp. 487-531.
Heim, I. (1992), "Presupposition projection and the semantics of attitude verbs", *Journal of Semantics*, 9, pp. 183-221.
Heim, I. (1993), "Anaphora and semantic interpretation: a reinterpretation of Reinhard's approach", Tübingen working papers.
Heim, I. (1994), "Comments on Abusch's Theory of Tense", In: H. Kamp (ed.), *Ellipsis, Tense and Questions, Dyana-2 Deliverable R.2.2.B.* Amsterdam.
Heim, I. (ms), "Presupposition projection in modal and attitude contexts", University of Texas, Austin. (an early version of Heim (1992))
Hendriks, H. and P. Dekker (1996), "Links without locations", In: P. Dekker and M. Stokhof (eds.), *The proceedings of the 10'th Amsterdam Colloquium*, Amsterdam.
Heusinger, K. von. (1995), "Reference and Salience", In: F. Hamm *et al.* (eds.) *The Blaubeuren Papers, Proceedings of the Workshop on Recent Developments in the Theory of Natural Language Semantics*, pp. 149- 172.
Hintikka, J. (1962), *Knowledge and Belief*, Ithaca, NY: Cornell University Press.
Hintikka, J. (1969), *Models for Modalities*, D. Reidel, Dordrecht.
Hintikka, J. (1975), *The Intentions of Intentionality*, D. Reidel, Dordrecht.
Hintikka, J. and J. Kulas, (1985), *Anaphora and Definite Descriptions, Two applications of Game-Theoretic Semantics*, Reidel, Dordrecht.
Horn, L. (1969), "A presuppositional aproach to *only* and *even*", *CLS*, 5, pp.98-107.
Jackendoff, R. (1972), *Semantic interpretation in generative grammar*, MIT Press, Cambridge.
Janssen, T. (1984), "Individual concepts are useful", In: F. Landman and F. Veldman (eds.): *Varieties of Formal Semantics*. Dordrecht, pp. 171–192.
Janssen, T. (1986), *Foundations and Applications of Montague Grammar*, Amsterdam, CWI.
Jeffrey, R. (1965), *The Logic of Decision*, McGraw-Hill, New York. University of Chicago Press, Chicago.
Jeffrey, R. and R. Stalnaker (1994), "Conditionals as random variables", In: E. Eells and B. Skyrms (eds), *Probability and conditionals*, Cambridge University Press, Cambridge.
Kadmon, N. (1990), "Uniqueness", *Linguistics and Philosophy*, 13, pp. 273-324.
Kamp, J.A.W. (1971), "Formal properties of 'Now'", *Theoria*, 37, pp. 227-273.
Kamp, J.A.W. (1973), "Free choice permission", *Proceedings of the Aristotelian Society*, N.S., 74, pp. 57-74.
Kamp, J.A.W. (1979), "Semantics versus pragmatics", In: F. Guenthner and J. Smidt (eds.), *Formal Semantics and Pragmatics for Natural Language*, Reidel, Dordrecht, pp. 225-78.
Kamp, J.A.W. (1981), "A theory of truth and semantic representation", In: Groenendijk *et al.* (eds), *Formal Methods in the Study of Language*, Amsterdam, pp. 277-322.
Kamp, J.A.W. (1985), "Context, thought and communication", *Proceedings of the Aristotelian Society*, 85, pp. 239-261.
Kamp, J.A.W. (1988), "Comments on Robert Stalnaker: 'Belief attribution and context'", In: R. Grimm and D. Merrill (eds.), *Contents of Thought*, Tuscon, University of Arizona Press.
Kamp, J.A.W. (1990), "Prolegomena to a structural account of belief and other attittudes", In: C. A. Anderson and J. Owens (eds.), *Propositional Attitudes, The role of Content in Logic, Language, and Mind*, CSLI Lecture Notes, Nr. 20, pp. 27-90.
Kamp, J.A.W. and A. Roßdeutcher (1992), "Remarks on lexical structure, DRS-construction and lexically driven inferences", *Arbeitspapiere des Sonderforschungsbereichs 340*, Stuttgart.

Kamp, J.A.W. and U. Reyle, (1993), *From Discourse to Logic*, Kluwer, Dordrecht.
Kaplan, D. (1969), "Quantifying in", In: D. Davidson and J. Hintikka (eds.), *Words and Objections, Essays on the work of W.V. Quine*, Dordrecht, pp. 178-214.
Kaplan, D. (1975), "How to Russell a Frege-Church", *The Journal of Philosophy*, IXXII, pp. 716-729.
Kaplan, D. (1978), "Dthat", In: P. Cole (ed.), *Syntax and Semantics, Vol. 9: Pragmatics*, New York, pp. 221-243.
Kaplan, D. (1989), "Demonstratives", In: I. Almog *et al.* (eds.), *Themes from Kaplan*, Oxford University Press, pp. 481-563.
Karttunen, L. (1969), "Pronouns and variables", In: R. I. Binnick *et al.* (eds), *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, Chicago.
Karttunen, L. (1973), "Presuppositions in compound sentences", *Linguistic Inquiry*, 4, pp.167-193.
Karttunen, L. (1974), "Presuppositions and linguistic context", *Theoretical Linguistics*, 1, pp.181-194.
Karttunen, L. (1977), "Syntax and semantics of questions", *Linguistics and Philosophy*, 1.1, pp. 3-44.
Karttunen, L. and S. Peters, (1979), "Conventional implicature", In: C.K. Oh and D. Dinneen (eds.), *Syntax and Semantics, vol 11: Presupposition*, Academic Press, New York, pp.1-56.
Katsuno, H. and A. Mendelzon, (1991) "On the difference between updating a knowledge database and revising it", In: *Proceedings of the Second International conference on Principles of Knowledge Representation and Reasoning*, pp. 387-394.
Katz, G. (ms.), *Entailment relations among intensional sentences*, University of Tübingen.
Kibble, R. (1994), "Dynamics of epistemic modality and anaphora", In: H. Bunt *et al.* (eds.), *International Worksop on Computational Semantics*, pp. 121- 130.
Krahmer E. (1995), *Discourse and Presupposition*, Ph.D. dissertation, University of Tilburg.
Krahmer, E. and R. Muskens (1995), "Negation and disjunction in discourse representation theory", *Journal of Semantics*, 12, pp. 357-376.
Kratzer, A. (1981), "Partition and revision: the semantics of counterfactuals", In: *Journal of Philosophical Logic*, 23, pp. 35-62.
Kratzer, A. (1989), "An investigation of the lumps of thought", *Linguistics and Philosophy*, 12, pp. 607-653.
Kratzer, A. (ms), *How specific is a fact*, University of Massachusetts, Amherst.
Kratzer, A. (ms.), *Scope or Pseudoscope? Are there wide scope indefinites?*, University of Massachusets, Amherst.
Kripke, S. (1963), "Semantic considerations in modal logic", *Acta Fennica*, 16, pp. 83-94.
Kripke, S. (1971), "Identity and necessity", In: M. Munitz (ed.), *Identity and Individuation*, New York University Press.
Kripke, S. (1972/80), "Naming and necessity", In: D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Dordrecht, pp. 253-355, 763-769.
Kripke, S. (1977), "Speakers reference and semantic reference" *Midwest Studies in Philosophy*, II, pp. 255-276.
Kripke, S. (1979), "A puzzle about belief", In: A. Margalit (ed.), *Meaning and Use*, Dordrecht, pp. 239-283.
Kripke, S. (ms.), *Presupposition and anaphora: Remarks on the formulation of the projection problem*, Princeton University.
Lakoff, J. (1972), "Linguistics and natural logic", In: D. Davidson and G. Harman (eds.) *Semantics of Natural Language*, Dordrecht: Reidel, pp. 545-665.
Landman, A. (1986), "Conflicting presuppositions and modal subordination", In: *Papers from the 22nd Regional Meeting, Chicago Linguistic Society*, pp. 195-207.
Leblanc, H. (1983), "Alternatives to standard first-order semantics", In: D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic, Vol. 1*, pp. 189 - 274.
Lerner, J.Y. and T.E. Zimmermann, (1983), "Presuppositions and quantifiers", In: Bäuerle *et al.* (eds.), *Meaning, Use, and Interpretation of Language*, Walter de Gruyter, Berlin

Lerner, J.Y. and T.E. Zimmermann, (1984), "Bedeutung und Inhalt von Eigennamen", Papier Nr. 94 des SFB 99. Konstanz.
Lewis, D.K. (1968), "Counterpart theory and quantified modal logic", *The Journal of Philosophy*, 65, pp. 113-126.
Lewis, D.K. (1970), "General semantics", *Synthese*, 22, pp. 18-67.
Lewis, D.K. (1973), *Counterfactuals*, Oxford: Blackwell.
Lewis, D.K. (1974), ""Radical interpretation", *Synthese*, 23, pp. 331- 344.
Lewis, D.K. (1975a), "Probabilities of conditionals and conditional probabilities", *The Philosophical Review*, 3, pp. 297-315.
Lewis, D.K. (1975b), "Adverbs of quantification", In: E. Keenan (ed.) *Formal Semantics of Natural Language*, Cambridge, UP, pp. 2-15.
Lewis, D.K. (1978), "Truth in fiction", *American Philosophical Quarterly*, 15, pp. 37-46.
Lewis, D.K. (1970/9), "A problem about permission", In: E. Saarinen *et al.* (eds.), *Essays in Honour of Jaakko Hintikka*, Reidel, Dordrecht (ms. 1970).
Lewis, D.K. (1979a), "Attitudes de dicto and de se", *Philosophical Review*, 88, pp. 513-543.
Lewis, D.K. (1979b), "Scorekeeping in a language game", *Journal of Philosophical Logic*, 8, pp. 339-359.
Lewis, D.K. (1979c), "Counterfactual dependence and times's arrows", *Nous*, 13, pp. 455-476.
Lewis, D.K. (1980), "Index, context and content", In: S. Kanger and S. Ohman (eds.) *Philosophy and Grammar*, pp. 79-100.
Lewis, D.K. (1981a), "Causal decision theory", *Australian Journal of Philosophy*, 59, pp. 5-30.
Lewis, D.K. (1981b), "What puzzling Pierre does not believe", *Australasian Journal of Philosophy*, 59, pp. 283-289.
Lewis, D.K. (1982), "'Whether' reports", In: T. Pauli *et al.* (eds.), *320311: Philosophical Essays Dedicated to Lennart Acvist on his Fiftieth Birstday*, Filosofiska Studier.
Lewis, D.K. (1983a), "Individuation by acquaintance and by stipulation", *Philosophical Review*, 92, pp. 3-12.
Lewis, D.K. (1983b), *Philosophical Papers, Vol I*, Oxford.
Lewis, D.K. (1986), *On the Plurality of Worlds*, Oxford.
Lewis, D.K. (ms.), *Elusive Knowledge*.
Loar, B. (1986), "Social content and psychological content", In: R. Grimm and D. Merrill (eds.), *Contents of Thought*, Tuscon, pp. 99-110.
Lyons, J. (1977), "Deixis and anaphora", In: T. Meyers (ed.), *The Development of Conversation and Discourse*, Edingburgh: Edingburgh University Press.
McGee, V. (1994), "Learning the impossible", In: E. Eells and B. Skyrms (eds), *Probability and Conditionals*, Cambridge University Press, Cambridge.
McKay, T. (1991), "Representing *de re* beliefs", *Linguistics and Philosophy*, 14, pp. 711-739.
McKay, T. and P. van Inwagen (1977), "Counterfactuals with disjunctive antecedents", *Philosophical studies*, 31, pp. 353-356.
Merin, A. (1992), "Permission sentences stand in the way of Boolean and other lattice-theoretic semantics", *Journal of Semantics*, 9, pp. 95-162.
Meyer Viol, W.M.P. (1995), *Instantial Logic, An investigation into reasoning with instances*, Ph.D. dissertation, ILLC Dissertation Series 1995-II, Amsterdam
Montague, R. (1974), *Formal Philosophy*, New Haven, Yale University Press.
Morreau, M. (1992) *Conditionals in Philosophy and Artificial Intelligence*, Ph.D. dissertation, University of Stuttgart.
Muskens, R. (1989), *Meaning and Partiality*, Ph.D. dissertation, University of Amsterdam.
Neale, S. (1990), *Descriptions*, MIT Press, Cambridge.
Nute, D. (1984), "Conditional logic", In: D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic, Vol II*, D. Reidel, Dordrecht.
Parsons, T. (1969), "Essentialism and quantified modal logic", *The Philosophical Review*, LXXVIII, 1, pp. 35-52.

Partee, B. (1972), "Opacity, coreference, and pronouns", In: D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Reidel, Dordrecht, pp. 415-441.
Pearl, J. (1994), "From Adams' conditionals to default expressions, causal conditionals, and counterfactuals", In: Eells and B. Skyrms (eds), *Probability and conditionals*, Cambridge University Press, Cambridge.
Perry, J, (1977), "Frege on demonstratives", *Philosophical Review*, 86, pp. 474-497.
Perry, J. (1979), The problem of the essential indexical", *Noûs*, 13, pp. 3-31.
Perry, J. (1980), "A problem about continued belief", *Pacific Philosophical Quarterly*, 61, pp. 317-332.
Peters, S. (1977), "A truthconditional formulation of Karttunen's account of presuppositions", In: *Texas Linguistic Forum 6*, Department of Linguistics, University of Texas at Austin, pp. 137-149.
Plantinga, A. (1974), *The Nature of Necessity*, Oxford University Press.
Popper, K.R. (1959), *The Logic of Scientific Discovery*, London, Hutchinson.
Portner, P.H. (1992), *Situation Theory and the Semantics of Propositional Expressions*, Ph.D. dissertation, University of Massachusetts, Amherst.
Powers, L. (1976), "Comments on 'Propositions'", In: A. Mackay and D. Merrill (eds.), *Issues in the Philosophy of Language*, New Haven, Yale Univeristy Press.
Price, H. (1989), "Defending desire-as-belief", *Mind*, pp. 119-127.
Putnam, H. (1975), "The meaning of 'meaning'", In: K. Gunderson (ed.), *Language, Mind and Knowledge*, Minneapolis, MN, University of Minnesota Press.
Quine, W.V. (1943), "Notes on existence and necessity", *Journal of Philosophy*, XL, pp. 113-127.
Quine, W.V. (1953/80), "Reference and modality", In: W.V. Quine, *From a Logical Point of View*, Harvard University Press.
Quine, W.V. (1956), "Quantifiers and propositional attitudes", *The Journal of Philosophy*, 53, pp. 177-187.
Quine, W.V. (1960), *Word and Object*, Cambridge, Mass.
Ramsey, F.P. (1931), "Truth and probability", In: R.B. Braithwaite (ed.), *The Foundations of Mathematics and other Logical Essays*, London, Rootledge and Kegan Paul.
Reinhard, T. (1992), "Wh-in-situ: an apparent paradox", In: P. Dekker and M. Stokhof (eds.), *Proceedings of the Eigth Amsterdam Colloquium*.
Rescher, N. (1967), "Semantic foundations for the logic of preference", In: N. Rescher (ed.), *The Logic of Decision and Action*, University of Pittsburgh Press.
Reyle, U. (1993), "Dealing with ambiguities by underspecification: construction, representation and deduction, *Journal of Semantics*, 10, 2, pp.
Richard, M. (1983), "Direct reference and ascription of belief", *Journal of Philosophical Logic*, 12, pp. 425-452.
Richard, M. (1993), "Attitudes in context", *Linguistics and Philosophy*, 16, pp. 123-148.
Roberts, C. (1989), "Modal subordination and pronominal anaphora in discourse", *Linguistics and Philosophy*, 12, pp. 683-721.
Rohrbaugh, G. (1996), "An event-based semantics for deontic utterances", In: P. Dekker and M. Stokhof (eds.), *Proceedings of the 10'th Amsterdam Coloquium*, Amsterdam.
Rooth, M. (1992), "A theory of focus interpretation", *Natural Language Semantics*, 1, pp. 75-116.
Rooy, van R.A.M. (1994), "A two-dimensional account of presuppositions in quantified contexts", In: F. Hamm *et al.* (eds.), *The Blaubeuren Papers, Proceedings of the Workshop on Recent Developments in the Theory of Natural Language Semantics*, pp. 305-346.
Rooy, van R.A.M. (1997), "Anaphoric pronouns as referential expressions", In: P. Weingartner *et al.* (eds.), *The Role of Pragmatics in Contemporary Philosophy: Contributions of the Austrian Ludwig Wittgenstein Society*, Vol. 5, Kirschberg.
Rooy, van R.A.M. (to appear), "Anaphoric relations across belief contexts", In: K. von Heusinger and U. Egli (eds.), *Reference and Anaphoric relations*, Kluwer.

Rooy, van R.A.M. and T.E. Zimmermann, (1996), "An externalist account of intentional identity", In: K. von Heusinger and U. Egli (eds.), *Proceedings of the Konstanz Workshop "Reference and Anaphoric Relations"*, Konstanz.
Russell, B. (1905), "On denoting", *Mind*, 14, pp. 479-493.
Russell, B. (1917), "Knowlegde by acquaintance and knowledge by description", In: B. Russell, *Mysticism and Logic*, London, pp. 152-167.
Russell, B. (1957), "Mr. Strawson on referring", *Mind*, 66, pp. 385-389.
Saarinen, E. (1978), "Intentional identity interpreted", *Linguistics and Philosophy*, 2, pp. 151-223.
Saebo, K.J. (1992), "Anaphoric presupposition and zero anaphora", In: H. Kamp (ed.), *Presupposition*, ILLC University of Amsterdam, Amsterdam, Dyana-2 deliverable R2.2.A, part II.
Salmon, N. (1986), *Frege's Puzzle*, M.I.T. Press, Bradford Books.
Salmon, N. (1987), "Reflexivity", *Notre Dame Journal of Formal Logic*, 27,pp. 401-429.
Sandt, R.A. van der (1982), *Kontekst en Presuppositie*, Ph.D. Dissertation, University of Nijmegen.
Sandt, R.A. van der (1988), *Context and Presupposition*, Croom Helm, London.
Sandt, R.A. van der (1989), "Presupposition and discourse structure", In: R. Bartsch *et al.* (eds.), *Semantics and Contextual Expression*, Foris, Dordrecht, pp. 287-294.
Sandt, R.A. van der (1991), "Denial", *Papers from the Parasession on Negation*, Chicago Linguistic Society, Chicago.
Sandt, R.A. van der (1992), "Presupposition projection as anaphora resolution", *Journal of Semantics*, 9, 4, pp. 223-267.
Seuren, P.A.M. (1985), *Discourse Semantics*, Oxford.
Schiffer, S. (1990), "The mode-of-presentation problem", In: C.A. Anderson and J. Owen, *Propositional Attitides, the role of Content in Logic, Language and Mind*, CSLI lecture notes, nr. 20, pp. 249-268.
Segerberg, K. (1973), "Two dimensional modal logic", *Journal of Philosophical Logic*, 2, pp. 77-96.
Shackle, G.I.S. (1961), *Decision, Order and Time in Human Affairs*, Cambrigde University Press, Cambrigde.
Skyrms, B. (1980a), *Causal Necessity*, Yale University Press, New Haven, Conn.
Skyrms, B. (1980b), "The prior prospensity account of subjunctive conditionals", In: W.L. Harper *et al.* (eds), *Ifs*, D. Reidel, Dordrecht.
Skyrms, B. (1994), "Adams conditionals", In: E. Eells and B. Skyrms (eds), *Probability and conditionals*, Cambridge University Press, Cambridge.
Slater, B.H. (1988), "Intensional identities", *Logique et Analyse*, 2, pp. 93-107.
Smaby, R. (1979), "Ambiguity of pronouns: A simple case", In: U. Mönnich (ed.), *Aspects of Philosophical Logic*, pp. 129-156.
Smullyan, A.F., (1948), "Modality and description", *Journal of Symbolic Logic*, 13, pp. 31-37.
Soames, S. (1979), "A projection problem for speaker presuppositions", *Linguistic Inquiry*, 10, pp. 623-666.
Soames, S. (1982), "How presuppositions are inherited: a solution to the projection problem", *Linguistic Inquiry*, 13.3, pp. 483-545.
Soames, S. (1987), "Direct reference, propositional attitudes, and semantic content", *Philosohical Topics*, 5, pp. 47-87.
Soames, S. (1989), "Presupposition", In: D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic, Vol. IV*, Dordrecht, pp. 553-616.
Soames, S. (1990), "Pronouns and propositional attitudes", *Proceedings of the Aristotelian Society*, pp. 191-212.
Sommers, F. (1982), *The Logic of Natural Language*, Oxford: Clarendon Press.
Spohn, W. (1975), "An analysis of Hansson's dyadic deontic logic", *Journal of Philosophical Logic*, 4, pp. 237-252.
Spohn, W. (1987), "Ordinal conditional functions: a dynamic theory of epistemic states", In: *Causation in Decision, Belief Change, and Statistics*, W.L. Harper and B. Skyrms (eds.), Dordrecht: Reidel, vol. 2, pp. 105-134.

Spohn, W. (1997), "Über die Gegenstande des Glaubens", In: G. Meggle and P. Steinacker, (eds.), *Proceedings of the 2nd Conference "Perspectives in Analytic Philosophy"*, de Gruyter.
Spohn, W. (ms), "Begründungen a prioi - oder: ein frischer Blick auf Dispositionsprädikate", University of Konstanz.
Stalnaker, R.C. (1968), "A theory of conditionals", *Studies in Logical Theory*, American Philosophical Quarterly Monograph Series, No. 2, Blackwell Oxford.
Stalnaker, R.C. (1970a), "Probability and conditionals", *Philosophy of Science*, 37.
Stalnaker, R.C. (1970b), "Pragmatics", *Synthese*, 22, pp. 272-289.
Stalnaker, R.C. (1972), "Propositions", In: A. Mackay and D. Merrill (eds.), *Issues in the Philosophy of Language*, New Haven and London, Yale Univeristy Press, pp. 197-213.
Stalnaker, R.C. (1973), "Presuppositions", *Journal of Philosophical Logic*, 2, pp.447-457.
Stalnaker, R.C. (1974), "Pragmatic presupposition", In: Munitz and Unger (eds.), *Semantics and Philosophy*, NYP.
Stalnaker, R.C. (1975), "Indicative conditionals", *Philosophia*, 5.
Stalnaker, R.C. (1976a), "Possible worlds", *Nous*, 10, pp. 65-75.
Stalnaker, R.C. (1976b), "Letter to Bas van Fraassen", In: W.L. Harper and C.A. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol. I, Reidel, Dordrecht.
Stalnaker, R.C. (1976c), "Letter to William Harper", In: W.L. Harper and C.A. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol. I, Reidel, Dordrecht.
Stalnaker, R.C. (1977), "Complex predicates", *The Monist*, 60, pp. 327-339.
Stalnaker, R.C. (1978), "Assertion", In: P. Cole (ed.), *Syntax and Semantics, vol. 9: Pragmatics*, pp. 315-332.
Stalnaker, R.C. (1979), "Anti-essentialism", In: P. Frech *et al.* (eds.), *Midwest studies in Philosophy, 4, Metaphysics*, pp. 343-355.
Stalnaker, R.C. (1980a), A defense of conditional excluded middle", In: *Ifs*, W.L. Harper *et al.* (eds.), Reidel, Dordrecht.
Stalnaker, R.C. (1980b), "Letter to David Lewis", In: *Ifs*, W.L. Harper *et al.* (eds.), Reidel, Dordrecht, (Mai 21, 1972).
Stalnaker, R.C. (1981), "Indexical belief", *Synthese*, 49, pp. 129-151.
Stalnaker, R.C. (1984), *Inquiry*, Cambridge, MA, MIT Press/Bradford Books.
Stalnaker, R.C. (1987a), "Semantics for belief", *Philosophical Topics*, 15, pp. 177-190.
Stalnaker, R.C. (1987b), "Counterparts and identity", *Midwest Studies in Philosophy Studies in Essentialism*, 11, pp. 121-140.
Stalnaker, R.C. (1988), "Belief attribution and context", In: R. Grimm and D. Merrill (eds.) , *Contents of Thought*, Tuscon, University of Arizona Press.
Stalnaker, R.C. (1989), "On what's in the head", *Philosophical Perspectives, 3: Philosophy of Mind and Action Theory.*
Stalnaker, R.C. (1990a), "Mental content and linguistic form", *Philosophical Studies*, 58, pp. 129-146.
Stalnaker, R.C. (1990b), "Narrow content", In: C.A. Anderson and J. Owens (eds.), *Propositional Attitudes: The Role of Content in Logic, Language and Mind*, Standford, CSLI.
Stalnaker, R.C. (1991), "The problem of logical omniscience, I", *Synthese*, 89, pp. 425-440.
Stalnaker, R.C. (1993), "Twin Earth revisited", *Proceedings of the Aristotelian Society*, XCIII, pp. 297-311.
Stalnaker, R.C. (1994), "Stalnaker", In: S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Blackwell Oxford, pp. 561-568.
Stalnaker, R.C. (1996), "On the representation of context", In: T. Galloway and J. Spence (eds.), Proceedings of Salt VI, Cornell University, Mass.
Stalnaker, R.C. (ms.), *Comments on Lewis's problem about permission.*

Stampe, D. W. (1977), "Towards a causal theory of linguistic representation", *Midwest Studies in Philosophy ,II: Studies in the Philosophy of Language*, pp. 42-63, Morris, Minn.: University of Minnesota at Morris.

Stechow, A. von (1984), "Structured propositions and essential indexicals", In: F. Landman and F. Veltman (eds.), *Varieties of Formal Semantics*, Dordrecht, pp. 385-403.

Stechow, A. von (1990), "Focusing and background operators", In: *Discourse Particles, Pragmatics and Beyond*, John Benjamins, Amsterdam.

Stechow, A von and T.E. Zimmermann (1984), "Term answers and contextual change", *Linguistics*, 22, pp. 3-40.

Strawson, P.F. (1950), "On referring", *Mind*, 59, pp. 320-344.

Strawson, P.F. (1952), *Introduction to Logical Theory*, London: Methuen.

Strawson, P.F. (1964), "Identifying reference and truth-values", *Theoria*, 30, pp. 96-118.

Tichy, P. (1976), "A counterexample to the Stalnaker-Lewis analysis of counterfactuals", *Philosophical sStudies*, 29, pp. 271-273.

Thijsse, E. (1992), *Partial Logic and Knowledge Representation*, Eburon Publishers, Delft.

Thomason, R.H. (1984), "Combinations of tense and modality", In: D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic, Vol. II*, D. Reidel, Dordrecht.

Thomason, R.H. and R.C. Stalnaker, (1968), "Modality and reference", *Nous*, pp. 359-372.

Thomason, R.H. and A. Gupta, (1980), "A theory of conditionals in the context of branching time", In: W.L. Harper *et al.* (eds.), *Ifs*, Reidel, Dordrecht.

Veltman, F. (1976), "Prejudices, presuppositions and the theory of counterfactuals", In: J. Groenendijk and M. Stokhof (eds), *Amsterdam Papers in Formal Grammar*, Amsterdam, vol.1.

Veltman, F. (1990), "Defaults in update semantics", In: Kamp (ed.), *Conditionals, Defaults and Belief Revision*, CCS, Edinburh,Dyana deliverable R2.5.A.

Warmbrod, K, (1981), "Counterfactuals and substitution of equivalent antecedents", *Journal of Philosophical Logic*, 10, pp. 267-289.

Weijters, T. (1989), *Denotation in Discourse, Analysis and Algorithm*, Ph.D. dissertation, University of Nijmegen.

Westerstahl, D. (1984), "Determiners and contexts sets", In: J. van Benthem and A ter Meulen (eds.), *Generalised Quantifiers in Natural Language*, Dordrecht, pp. 45-71.

Wilson, D. (1975), *Presuppositions and Non-Truth-Conditional Semantics*, New York, Academic Press.

Winter, Y. (1996), "Choice functions and the scopal semantics of indefinites", In: K. von Heusinger and U. Egli (eds.), *Proceedings of the Konstanz Workshop "Reference and Anaphoric Relations"*, Konstanz.

Wright, H. von (1963), *The Logic of Preference*, Edinburgh.

Zeevat, H. (1987), "A treatment of belief sentences in discourse representation theory", In: J. Groenendijk *et al.* (Eds.), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, Dordrecht, Foris Publications.

Zeevat, H. (1992), "Presupposition and accommodation in update semantics", *Journal of Semantics*, 9, pp. 379-412.

Zeevat, H. (1996), "A neoclassical analysis of belief sentences", In: P. Dekker and M. Stokhof (eds.), *Proceedings of the 10'th Amsterdam Coloquium*, Amsterdam.

Zimmermann, T.E. (1991), "Kontextabhangigkeit", In: A. von Stechow and D. Wunderlich (eds.), *Semantik: ein internationales Handbuch der zeitgenossischen Forschung*, Berlin, pp. 156-229.

Zimmermann, T.E. (1995), "Subjects in perspective. Comments on Paul Dekker's 'On context and identity'", In: H. Kamp and B. Partee (eds.), *Context in the Analysis of Linguistic Meaning*, Stuttgart/Prague.

Zimmermann, T.E. (forthcomming), "Remarks on the epistemic role of discourse referents", In: L. Moss (ed.), *Proceedings of ITALLC 96*, London.

Zimmermann, T.E. (ms.), "Do we bear attitudes towards quantifiers?", University of Stuttgart.

Zusammenfassung

Der durchgehende Leitgedanke dieser Dissertation dient der Verteidigung der Ansicht, daß Bedeutung und Inhalt linguistischen Ausdrucks vor dem Hintergrund der Intentionen, des Glaubens und der Konventionen der Benutzer der Sprache erklärt werden sollten, und daß der Inhalt dessen, was intendiert und geglaubt oder präsupponiert wird, teilweise im Bezug auf Kaussale und externe Faktoren erklärt werden sollte. Nur wenn wir annehmen, daß der Inhalt unserer Einstellungen durch externalistische Begriffe erklärbar ist, können wir unsere Behauptungen mit einer Kausale Referenztheorie vereinbaren, für die Kripke und andere überzeugende Argumente geliefert haben. Das heißt, nur so können wir diese Kausale Referenztheorie mit der Intuition vereinbaren, daß Referenz vor dem Hintergrund dessen, welche Handlung Sprecher durch den Gebrauch eines Begriffes vornehmen, erklärt werden sollte, und nicht durch Eigenschaften des Begriffes selbst.

Im ersten Kapitel (*Belief and Belief Attribution*) diskutiere ich, wie Glaubenszustände repräsentiert werden müssen, damit wir unsere semantischen Intuitionen rechtfertigen können, welche die Glaubenszuschreibung betreffen. Der nächstliegenden Deutung von Glaubenszuschreibungen zufolge ergibt sich die Bedeutung eines Satzes "a glaubt, daß A" in einem gegebenen Kontext *c* kompositionell aus den Bedeutungen seiner Teile in *c*. Danach scheint es, als erzwängen gewisse Ersetzungsprobleme die Annahme, daß Bedeutungen und Inhalte ausgesprochen feinstrukturierte Gebilde seien und daß Glaubenszustände auf sehr feinkörnige Weise modelliert werden müssen. Daß Problem dieser Methode ist, daß man kaum *unabhängige* Gründe für eine derart feine Strukturierung des Inhalts finden wird. Die Alternative wäre eine Strategie, die von vornherein, unabhängig von der Glaubenszuschreibung, eine philosophisch motivierte Vorstellung von Inhalt präsentiert. Diese unabhängige Vorstellung wäre zwangsläufig eine relativ grob strukturierte. Der Leitgedanke des ersten Kapitels dieser Dissertation ist die Verteidigung dieser Strategie und dessen, was ich als deren Konsequenz ansehe: Aus der Sicht dessen, der etwas glaubt, sollte der Glaubenszustand mit einer Proposition, die durch Wahrheitsbedingungen entsteht, modelliert werden. In diesem Fall sollte es voraussehbar sein, daß Glaubensinhalte wechselseitig konsistent und unter logischer Implikation abgeschlossen sind. Dieses Bild werde ich verteidigen, desgleichen seine Konsequenzen, indem ich für eine drei-Wege-Strategie zur Lösung einiger Probleme, die bei diesem Ansatz entstehen, argumentiere. Die Probleme wären, zu erklären, (i) warum so viele Glaubenszuschreibungen zutreffend und wahr sind, (ii) warum sie sich meistens auf tatsächliche Referenten beziehen und (iii) warum wir doch nicht folgern, daß Subjecten in einem intern inkonsistenten Glaubenszustand sind. Auf diese Probleme wird gemäß meiner Alternativstrategie eingegangen. Die Erklärung beruht teils darauf, daß die durch den eingebetteten Satz ausgedrückte Proposition das Verhältnis zwischen der Repräsentation und ihrem Inhalt betrifft, teils auf der Intention und Präsupposition des Subjecten, der die Zuschreibung vornimmt, und teils darauf, daß Subjecten auf verschiedene Weise zu dem selben Objekt in kausaler Beziehung stehen.

Der erste Schritt dieser Strategie ist, den externen Umständen, von denen das Denotat eines Ausdrucks abhängt, genügend Gewicht zu geben. Der Begriff *Wasser* beispielsweise bedeutet H_2O, weil wir damit normalerweise einen Stoff bezeichnen, der bestimmte oberflächlich zu beobachtende Eigenschaften hat, und es ist in unserer Welt einzig und allein H_2O, welches normalerweise und unter idealen Bedingungen diese Eigenschaften hat. Normalität und Idealbedingungen unterliegen jedoch Zufälligkeiten, es hängt von äußeren Gegebenheiten ab, welche Bedingungen für Normalität oder Optimalität gelten. In Putnams (1975) Geschichte von der Zwillingserde beispielsweise unterscheiden sich die oberflächlich beobachtbaren Eigenschaften des Stoffes, die auf der Zwillingserde mit *Wasser* assoziiert werden, von denen auf der Erde. Das heißt, die Normalitätsbedingungen bezüglich des Worts *Wasser* sind auf der Zwillingserde nicht die gleichen wie auf der Erde. Also

denotieren Zwillingserdensprecher mit ihrem Gebrauch des Wortes Wasser einen anderen Stoff als die Menschen auf der Erde. Auf die gleiche Weise hängt von relevanten Normalitätsbedingungen ab, was durch den eingebetteten Satz einer Glaubenszuschreibung ausgedrückt wird. Normalerweise bestimmen sich die Normalitätsbedingungen nach dem, was für uns normal ist; aber manchmal legen wir diese Normalitätsbedinungen auch danach fest, was für das Subjekt normal ist, dem wir den Glauben zuschreiben. Wenn der Subject mit einem bestimmten Begriff einen anderen Wert assoziiert als wir, und wenn wir bei der Festlegung der Normalitätsbedingungen den für ihn üblichen Gebrauch eines Begriffes berücksichtigen, sind wir in der Lage, den Glaubenszustand eines Subjectes, der, obwohl sie tatsächlich wahr ist, nicht glaubt, daß *Hesperus ist Phosphorus.* eine wahre Proposition denotiere, zu repräsentieren, ohne die Annahme aufzugeben, daß das Denotat eines Eigennamens durch die Kausale Referenztheorie determiniert sein sollte. Der Grund dafür ist, daß zur Erklärung der Glaubenszuschreibung die mit dem Glauben kompatiblen Welten nicht nur zur Bestimmung, ob eine Proposition wahr oder falsch ist, sondern auch, welche Proposition durch den Satz ausgedrückt wird, dienen. Dies wird in der formalen Darstellung durch zweidimensionale Modallogik deutlich, wo Welten diese zwei Rollen spielen, weil sie zwei Arten von Information enthalten: (i) Information über den *Gegenstand des Denkens* und (ii) Information über die linguistischen und sprachlichen Konventionen der Sprecher. Wenn Hans nicht glaubt, daß "Hesperus ist Phosphorus" eine (notwendig) wahre Proposition ausdrückt, glaubt er nicht, daß die Sprachkonventionen so sind, daß "Hesperus" und "Phosphorus" letzlich auf dasselbe Objekt verweisen.

Wir müssen das, was durch den eingebetteten Satz einer Glaubenszuschreibung ausgedrückt wird, davon abhängig machen, daß die Intentionen und Präsuppositionen derjenigen, welche die Zuschreibung vornehmen, geeignet sind, die Tatsache zu erklären, daß in den meisten konversationellen Kontexten wahrheitsgemäß und zutreffend behauptet werden kann: *Oskar glaubt daß Wasser das beste Getränk zum Löschen des Durstes ist*, so daß die Zuschreibung zwar H_2O betrifft, aber trotzdem nur von Oskars Standpunkt aus gemacht wird. Die Glaubenszuschreibung ist wahr, wenn Oskar in der Lage ist, die Alternativen, wo Wass er in unserer Einschätzung das beste Getränk zur Löschung des Durstes ist, von denen, wo es dies nicht ist, zu unterscheiden, und nur die ersteren als wahr eingeschätzt werden. Erst wenn der Skeptiker Fragen über das Wissen, oder wenn Putnam kritische Fragen über Bedeutungen stellt, geben wir das Präsupponierte auf und ziehen Alternativen in Betracht. Damit also die Glaubenszuschreibung wahr ist, verlangen wir mehr von dem Subject, wobei in der neuen konversationellen Situation die Zuschreibung sich als falsch herausstellen kann.

In Quines Geschichte scheint es uns gestattet zu sagen, daß Ralph glaubt daß Ortcutt ein Spion ist, weil er einen Mann mit braunem Hut gesehen hat, der sich verdächtig wie ein Spion benahm; außerdem wissen wir, daß dieser Mann Ortcutt war. Ebenso scheint es uns erlaubt zu sagen, daß Ralph nicht glaubt, daß Ortcutt ein Spion ist, weil Ralph am Strand einen Mann gesehen hat, der wie eine Stütze der Gesellschaft aussah, und er kommt überhaupt gar nicht auf die Idee, daß dieser Mann ein Spion sein könnte, obwohl dieser Mann tatsächlich auch Ortcutt war. Es ist offensichtlich, daß in diesem Fall Ralph aufgrund seiner internen repräsentationalen Mechanismen, die für Fakten über Ortcutt empfänglich sind, etwas über ihn glaubt. Die Schwierigkeit ist, daß er zwei Repräsentationen Ortcutts hat: er ist mit ihm auf zwei verschiedene Weisen bekannt. Also können wir wahrheitsgemäß sagen, daß in Quines Geschichte Ralph sowohl (1) *glaubt, daß Ortcutt ein Spion ist*, als auch (2) *nicht glaubt, daß Ortcutt ein Spion ist*. Wie ist dies ohne die Implikation, daß sein Glaubenszustand intern inkonsistent sei, möglich? Ich will argumentieren, daß wir diese Intuitionen am besten durch eine Counterpart-theorie erklären können. Diese läßt zu, daß ein Individuum der wirklichen Welt in den Welten, die eines Subjectes Glaubenszustand zu charakterisieren helfen, mehrere Repräsentanten hat.

Am Ende von Kapitel 1 entwickle ich ein System der Modallogik, welches sowohl die zweidimensionale Theorie von Kaplan (1989) wie Stalnaker (1978) umsetzt, und eine Version einer Counterpart-theorie. Die in Kapitel 1 entwickelten Ideen, die hinter letzterer Theorie stehen, sind in mehrfacher Hinsicht ungewöhnlich. Zunächst ist es eine Counterpart-theorie, die Intuitionen bezüglich zufälliger Identität bestätigt, obwohl es nur ein einziges Identitätsymbol gibt, welches den Gesetzen von Leibniz gehorcht. Dann werden *de re* Glaubenszuschreibungen erklärt, indem der Glaubenszustand eines Subjectes in einer Welt *w* durch das Paar <r,K> , *K* die Menge möglicher Welten, *r* die partielle Funktion, welche einem Individuum in *w* und einer Welt in *K* die Menge der Repräsentanten des Individuums in dieser Welt zuordnet, repräsentiert ist. So kann ich nämlich die Intuition rechtfertigen, daß Ralph glaubt und auch nicht glaubt, daß Ortcutt ein Spion ist, ohne zu implizieren, daß er sich in einem intern inkonsistenten Glaubenszustand befindet. Desgleichen wird die Intuition bestätigt, daß wenn Hans glaubt, daß Maria geht, er nicht von Willi glauben muß, daß dieser entweder spricht oder nicht spricht, wenn er *über* ihn gar nichts glaubt. Schließlich sagt das von mir vorgeschlagene System korrekterweise nicht voraus, daß wir (b) *John believes that Hesperus is Selfoutweighing*. aus (a) *John believes that Hesperus ouweighs Phosphorus.* folgern können, selbst wenn die Glaubenszuschreibung die tatsächliche Referenz der Bezeichnungen *Hesperus* und *Phosphorus* betrifft.

Dynamische semantische Theorien wie die Diskurs-Repräsentationstheorie von Kamp (1981) und die File-Change-Semantik von Heim (1982) erklären sehr treffend anaphorische Abhängigkeiten über Satzgrenzen hinweg. In diesen Theorien ist die Existenz von Diskursreferenten von wesentlicher Bedeutung. Deswegen enthalten hier die Informationszustände mehr als nur Wahrheitsbedingungen. Anstelle von Mengen möglicher Welten modellieren Mengen von Welt-Belegungsungspaaren die Informationszustände. Somit stellt sich aber die Frage, worin dieser zusätzliche Inhalt besteht und wofür die Diskursreferenten stehen. Der zusätzliche Inhalt scheint Information über den Diskurs selbst zu sein. Im zweiten Kapitel argumentiere ich, daß ein Diskursreferent für den vorausgesetzten *Sprecherreferenten* einer vom Sprecher verwendeten indefiniten Beschreibung steht, so daß anaphorische Pronomen normalerweise referentiell gebraucht werden müssen. Dies deckt sich mit dem pragmatischen Referenzkonzept, nach dem Referenz durch den Gebrauch eines Ausdrucks von den Sprechern bewirkt wird, nicht durch den Ausdruck selbst.

Angenommen, daß Pronomen normalerweise referentiell gebraucht werden, müssen wir zugestehen, daß der Hörer unter Umständen nicht erkennt, auf welches Objekt das Pronomen sich bezieht. Nehmen wir an, daß Pronomen normalerweise referentiell gebraucht sind, müssen wir den Gedanken aufgeben, daß der Hörer immer bestimmen kann, welche horizontale Proposition der Sprecher bei seiner Anwendung eines Satzes ausdrückt. Die Frage ist nur, ob wir diesen Gedanken wirklich aufgeben können. Ich behaupte, wir können und sollten es auch, denn in vielen Fällen weiß der Hörer nicht, welche horizontale Proposition der Sprecher ausdrückt. Der Grund dafür ist, daß für den Austausch von Informationen normalerweise nicht die horizontale Proposition zählt. Wie aber kann Kommunikation erfolgreich sein, wenn der Hörer nicht bestimmen kann, auf welches Objekt der Sprecher sich mit einem Indefinitum oder Pronomen beziehen wollte?

Zwar wird im Idealfall ein referentieller Ausdruck nur gebraucht, wenn der Hörer den Bezug erkennen kann, doch ist es klar, daß Idealbedingungen nicht immer zu erreichen sind. Wenn der Hörer etwas, das der Sprecher sagt, nicht zustimmt, kann es dafür zwei Gründe geben. Erstens kann bei präzisem Verständnis des Gesagten der Hörer eine Aussage über im Diskurs dargestellte Fakten ablehnen. Zweitens kann zwar Einverständnis zwischen Sprecher und Hörer vorhanden sein, der Hörer hat aber eine Aussage anders verstanden als sie vom Sprecher intendiert war. Letzteres mag vorkommen, wenn der Sprecher referentielle

Ausdrücke verwendet. Diese zwei unterschiedlichen Gründe für die Zurückweisung einer Aussage können im Rahmen von Kaplans (1989) zweidimensionaler Theorie der Referenz mittels der Stalnaker'schen (1978) *Diagonalisierungsstrategie* erklärt werden, da in dieser Theorie eine Unterscheidung zwischen zwei Arten von Tatsachen getroffen wird: (i) Tatsachen den Konversationsinhalt betreffend, (ii) Tatsachen, die in Sprachkonventionen und die Konversationssituation selbst eingebettet sind. Referenzkontexte repräsentieren Tatsachen der Konversationssituation, während Indizes Tatsachen des Konversationsinhalts repräsentieren. In einer zweidimensionalen Referenztheorie kann die Aussage eines Satzes durch eine Funktion von Referenzkontexten zur ausgedrückten Proposition dargestellt werden. Formal ist diese Funktion vom Referenzkontext zur Proposition natürlich ein Kaplan'scher (1989) *Charakter* oder ein Stalnaker'sches (1978) *Propositions Konzept*.

Die Information, die ein Hörer in einer Konversation erhält, ist jedoch normalerweise nicht exakt die Proposition, die im tatsächlichen Referenzkontext ausgedrückt wird. Der Grund hierfür liegt darin, daß der Kontext nicht nur die erreichbare Information für die Interpretation der kontextabhängigen Äußerungen repräsentiert, sondern auch die Informationen, welche von Sprecher und Hörer bezüglich des Konversationsinhaltes akzeptiert werden. Um diese zwei Arten von Information zu kombinieren, wird hier ein Kontext C durch eine Menge von Referenzkontext/Index-Paaren repräsentiert, in der alles, was in der Konversation akzeptiert wird gleichzeitig wahr ist. Sobald irgendein Element von C nach Einschätzung des Hörers das aktuelle Referenzkontext/indexpaar sein kann, kann er seinen Informationszustand aktualisieren, nachdem er die Äußerung akzeptiert und jedes Referenzkontext/indexpaar <c,w> in C, wo, was in *c* ausgedrückt wird, in *w* falsch ist, eliminiert hat. Dieser neue Informationszustand ist $\{<c,w> \in C \mid w \in [A](c)\}$. Er ist, was Stalnaker (1978) *die Diagonale von A in Bezug auf C* genannt hat. Normalerweise spiegelt dies wieder, was der Hörer ohne zu wissen, was der aktuelle Referenzkontext ist, empfangen kann. Normalerweise ist dies auch alles, was der Hörer zum erfolgreichen Informationsaustausch empfangen muß. Dennoch, so behaupte ich, ist es notwendig, das gesamte Propositions Konzept eines Satzes bezüglich seines Interpretationskontextes C zu bestimmen: Zur Erklärung der referentiellen Disambiguierung, zur Erklärung von Fragen und um eine Unterscheidung zwischen referentiellem und attributivem Gebrauch definiter Beschreibungen zu treffen.

Ich postuliere, daß der Gebrauch anaphorischer Pronomen meist als referentiell angesehen werden sollte, da das Pronomen einen Referenten eines relevanten Sprechers aufnimmt. Es ist wesentlich, daß der Sprecher so gehandelt hat, daß eines Sprechers Referent in den Diskurs eingeführt wurde, andernfalls kann das Pronomen nicht interpretiert werden. Normalerweise ist der relevante Akt, durch den der Referent des Sprechers eingeführt wird, des Sprechers Gebrauch eines Indefinitums. So erklärt sich der Gegensatz zwischen (a) *Hans hat einen Esel. Maria schlägt diesen.* und (b) **Hans ist ein Eselbesitzer. Maria schlägt diesen.* Anders als im letzteren Beispiel ist im ersteren das Indefinitum *ein Esel* verwendet, so daß dieses einen Sprecherreferenten hat. Das Ganze gründet zum einen darauf, daß ein Interpretationskontext die Diskursinformation enthält, welche Sätze von wem geäußert werden, zum anderen, daß es eine Übereinkunft der Sprecher ist, daß bei Gebrauch eines referentiellen Indefinitums ein bestimmtes Individuum gemeint ist.

Es ist offensichtlich, daß in einem Eselssatz wie *Wenn Hans einen Esel hat, schlägt er ihn.* das Indefinitum *ein Esel* keinen Sprecherreferenten hat. Um die anaphorische Abhängigkeit des Pronomens *ihn* von *ein Esel* zu erklären, schlage ich vor, daß auch Pronomen in Eselsätzen als referentielle Pronomen angesehen werden sollten, daß jedoch zur Bestimmung, welches Objekt durch deren anaphorische Antezedenten eingeführt ist, wir nicht den aktuellen Referenzkontext betrachten sollten, sondern daß alle hypothetischen oder

kontrafaktischen Referenzkontexte (oder Welten), in denen das Indefinitum einen Sprecherreferenten hatte, berücksichtigt werden sollten, die hinsichtlich des Subjekts der Konversation dieselben Tatsachen wahr machen wie die tatsächliche Welt.

Die in Kapitel 2 vorgeschlagene Theorie kommt den von Kamp (1981) und Heim (1982) vorgeschlagenen dynamischen semantischen Theorien sehr nahe. Es gibt jedoch mindestens drei wichtige Unterschiede: Erstens ist in den erwähnten Theorien der Status von Diskursreferenten nicht eindeutig, während in der von mir gerade skizzierten Theorie ein Diskursreferent die Information repräsentiert, die ein Hörer über einen bestimmten Sprecherreferenten hat. Zweitens sagt meine Theorie nicht vorher, daß der Diskurs *Ein Mann spaziert im Park umher. Er pfeift.* immer hinsichtlich der Wahrheitsbedingungen identisch ist mit *Ein Mann, der im Park umherspaziert, pfeift.*, während dies die Vorhersage der dynamischen Standardtheorien ist. Die zwei sind nicht als äquivalent vorherzusagen, wenn der Sprecherreferent des Indefinitums nicht pfeift, obwohl ein anderer Mann im Park dies tut. Drittens und letztens erlaubt der in dieser Dissertation skizzierte zweidimensionale Ansatz auf natürliche Weise zwei verschiedene Verwendungsweisen anaphorischer Pronomen. Entweder ist ein Pronomen referentiell verwendet und verweist auf den präsupponierten Sprecherreferenten seines indefiniten Antezedenten, oder es ist deskriptiv verwendet und verweist auf das einzige Objekt, das aus dem Satz, in dem sein syntaktischer Antezedent auftritt, erlangt werden kann (cf. Evans 1977). Nur wenn ein Pronomen deskriptiv verwendet wird, braucht es nicht auf den Sprecherreferenten seines syntaktischen Antezedenten zu verweisen, sofern es diesen überhaupt hat. Dies gründet darauf, daß die Regel, die den Referenten der Beschreibung determiniert, Teil der (horizontalen) Proposition ist, also unabhängig vom relevanten Referenzkontext. Indem wir deskriptive Pronomen im Rahmen einer dynamischen Theorie der Bedeutung einführen und umsetzen, können wir viele Phänomene erklären, die für herkömmliche dynamische Theorien problematisch sind.

So ist beispielsweise Partees Badezimmer-Beispiel *Entweder ist kein Badezimmer im Haus, oder es ist an einem ungewöhnlichen Ort.* nicht nur für die beliebten dynamischen Theorien der Anaphern problematisch, das Pronomen *es* kann in dem von mir vorgeschlagenen Rahmen ebensowenig als referentiell behandelt werden. Nach meiner Argumentation sollte es dies auch nicht, denn es ist ein dekriptives Pronomen, das stellvertretend für *das Badezimmer im Haus* steht. Wäre präsupponiert, daß es entweder keines oder mehr als ein Badezimmer im Haus gäbe, wäre der Gebrauch von *es* im zweiten Teilsatz unangemessen. Gleichgültig, ob ein Pronomen im Singular referentiell oder deskriptiv gebraucht wird, immer ist die Annahme von Einzigartigkeit involviert. Entweder wird das Pronomen referentiell gebraucht und verweist auf den einzigen, möglichen und relevanten, Sprecherreferenten, oder es wird deskriptiv verwendet und verweist auf das einzige Individuum in der möglichen Welt, welches der assoziierten Beschreibung genügt. Wegen der Annahme der Einzigartigkeit kann ich auch die Akzeptabilitätsdifferenz zwischen dem einwandfreien *Es stimmt nicht, daß keine Braut auf der Hochzeit ist. Sie steht direkt hinter dir.* und dem strukturell ähnlichen aber inakzeptablen (oder gewollt witzigen) *Es stimmt nicht, daß kein Gast auf der Hochzeit ist. Er steht direkt hinter dir.* erklären. Die Asymmetrie erklärt sich daraus, daß die Pronomen als deskriptive Pronomen behandelt werden müssen und es nur wahrscheinlich ist, daß es nicht mehr als eine Braut auf der Hochzeit gibt, nicht dagegen, daß es nur einen Gast gibt.

Wenn wir annehmen, daß manche Pronomen als deskriptive Pronomen behandelt werden sollten, ist es nur natürlich, andere Pronomen als *funktionale Pronomen* anzusehen. Dies ist in der Tat, was ich vorschlage und in dynamische Semantik umsetze. So erkläre ich Karttunens berühmtes Lohnscheck-Beispiel *Ein Mann, der seinen Lohnscheck seiner Ehefrau gibt, ist klüger als ein Mann, der ihn seiner Mätresse gibt.*, indem ich argumentiere, das Pronomen *ihn* verweise auf das Objekt, welches das Ergebnis einer Funktion von

Individuen zu ihren Lohnschecks sei, die auf den Mann im zweiten Teilsatz angewandt werde.

Zur Vorbereitung der Grundlage für die zweidimensionale Analyse von Präsuppositionen in quantifizierten Kontexten, welche ich im vierten Kapitel dieser Dissertation vorschlage, unterstütze und elaboriere ich den Gedanken, daß die Kontextabhängigkeit von Quantoren durch die Behandlung als Anaphern erklärt werden sollte. Dazu kommt die Idee, daß die für die Referenz in Frage kommenden Objekte nach Salienz geordnet werden sollten.

Das dritten Kapitel betrifft Glaubenszuschreibungen übergreifende anaphorische Bezüge. Alle semantischen Erklärungen, die Anaphern durch Variablen repräsentieren, haben das Problem, daß ein Pronomen im eingebetteten Teil einer Einstellungszuschreibung ein Indefinitum im eingebetteten Teil einer früheren Einstellungszuschreibung als Antezedenten haben kann. Dies ist in einer logischen Sprache zwar nicht schwer darzustellen, wenn das Indefinitum *de re* interpretiert wird, allein, daß dies nicht immer möglich ist, stellt das Problem dar, wie in Geachs berühmten Hob-Nob-Satz exemplifiziert ist: *Hob glaubt, daß eine Hexe Bobs Stute vergiftete, und Nob glaubt, daß sie Cobs Sau tötete.* Hier hilft die Annahme der Interpretierbarkeit mancher Pronomen als deskriptiv nicht wirklich weiter, da Nob nicht glauben muß, daß die Hexe, die er meint, Bobs Stute vergiftete, noch muß er glauben, daß Hob dies glaubt. Edelberg (1992) schlägt vor, daß Indefinita manchmal auf Glaubensobjekte von Subjecten verweisen und daß intentionale Identitätszuschreibungen zutreffen, wenn das Glaubensobjekt des ersten Subjecten (Hob), auf welches das Indefinitum verweist, ein Glaubensobjekt eines zweiten Subjecten (Nob) als Counterpart hat. In Kapitel 3 der Dissertation wird dieser Vorschlag präzisiert und in dynamische Semantik umgesetzt. Es wird argumentiert, daß (i) Glaubenszustände parallel zu Informationszuständen in der dynamischen Semantik modelliert werden können, daß (ii) auf diese Weise Glaubensobjekte als Information angesehen werden können, die solch ein Zustand über einen bestimmten Diskursreferenten oder den Wert einer Variablen hat, daß (iii) Indefinita in Glaubenskontexten auf Glaubensobjekte des relevanten Subjecten verweisen, und daß schließlich (iv) intentionale Identität in Form von Counterpartrelationen erklärt werden sollte. Edelberg (1992) argumentiert, daß zwei Glaubensobjekte verschiedener Subjecten counterparts von einander sind, wenn sie ungefähr die gleiche explanatorische Rolle in ihren entsprechenden Glaubenszuständen spielen. Im dritten Kapitel dieser Dissertation wird dagegen argumentiert, daß Beschränkungen für Counterpartrelationen nicht auf die Rollenähnlichkeit bezogen, sondern basierend auf der Kausalität dieser Glaubensobjekte formuliert werden sollten. Die Intuition ist, daß zwei Glaubensobjekte einander entsprechen, wenn sie die gleiche Quelle repräsentieren, und die gemeinsame Quelle ist normalerweise das Objekt, über das sie etwas glauben. Auf diese Weise wird auch eine plausible Analyse gewöhnlicher *de re* Glaubenszuschreibungen gegeben. De facto kann die in Kapitel 3 vorgeschlagene Analyse als Generalisierung der *de re* Glaubenszuschreibungen am Ende des ersten Kapitels angesehen werden. Die Relevanz der Theorien über Diskursrepräsentation und dynamische Semantik, wie sie im dritten Kapitel für die Analyse intentionaler Identität diskutiert werden, besteht in der Tatsache, daß in der vorgeschlagenen Analyse solcher Zuschreibungen *Glaubensobjekte* unentbehrlich betrachtet werden. Sie werden als partielle Objekte modelliert, die im Sinne von partielle Information über Variablenwerte in einer Art Informationszustand, wie er in Diskursrepräsentationstheorien entwickelt wurde, konstruiert sind. Auch um Edelbergs (1985) *Asymmetrieproblem* zu erklären, sollte nach meiner Argumentation dem begriff der Sprecherreferenz, welcher in Kapitel 2 entwickelt wird, entsprechendes Gewicht beigemessen werden.

Kapitel 4 konzentriert sich auf Präsuppositionen. Eine Präsupposition eines Satzes wird normalerweise als folgerung, die unter Negation erhalten bleibt, charakterisiert. So können wir sowohl aus *Hans bedauert, daß er durchgefallen ist.* wie aus *Hans bedauert nicht, daß er*

durchgefallen ist. schließen, daß Hans durchgefallen ist, also sagt man, diese Inferenz sei von Natur aus präsuppositional. Manchmal werden Präsuppositionsverhältnisse semantisch erklärt, Ich, aber, argumentiere, daß Präsupposition eine propositionale Einstellung ist, und daß das Präsupponierte dem vom Sprecher vorausgesetzten Allgemeinwissen zwischen Sprecher und Hörer entspricht. Also sollten Präsuppositionen pragmatisch analysiert werden: Was Sätze präsupponieren, sollte im Sinne dessen erklärt werden, was Sprecher normalerweise präsupponieren, wenn sie diese Sätze gebrauchen. Diese Analyse wird in Kapitel 4 zur Erklärung sogenannter präsuppositionaler Inferenzen vom Gebrauch bestimmter Sätze verwendet.

In den siebziger Jahren sah man die Semantik der Wahrheitsbedingungen als einen abgeschlossenen Themenkomplex an, der losgelöst von der Pragmatik studiert werden könne. Man argumentierte, daß, was normalerweise beim Gebrauch eines Satzes präsupponiert werde, vom *Inhalt* getrennt betrachtet werden *könne*, nämlich seine Wahrheitsbedingungen getrennt von der Aussage des Satzes. Darüber hinaus sah man es als *nützlich* an, Präsuppositionen von Wahrheitsbedingungen zu trennen, da man so die Frage des Inhalts von jener der Präsupposition trennen und somit die semantische Theorie vereinfachen konnte. Diese Sichtweise verhalf den *zweidimensionalen* Ansätzen für Präsuppositionen zur Geburt, denen zufolge der Behauptungsinhalt eines Satzes unabhängig von der Präsupposition bestimmt werden könne. Die bekanntesten zweidimensionalen Analysen von Präsuppositionen wurden von Gazdar (1979) beziehungsweise Karttunen & Peters (1979) entwickelt. Sie nahmen an, daß Behauptung und Präsupposition durch separate Propositionen repräsentiert werden könnten, die Behauptung unabhängig von der Präsupposition sei. Seit den achtziger Jahren werden zweidimensionale Erklärungen jedoch von fast allen Autoren abgelehnt. Die Gründe waren zum einen die große Aufmerksamkeit, welche die dynamischen Bedeutungstheorien dem Problem der Kontextabhängigkeit geschenkt hatten, zum anderen das *Bindungsproblem* in Karttunen & Peters (1979). Beide Probleme führten zu einem ähnlichen Schluß: Sind Präsupposition und Behauptung getrennt repräsentiert, wird die Abhängigkeit der Behauptung von der Präsupposition vernachlässigt.

Im Kapitel 4 stelle ich zwei Behauptungen auf und verteidige sie: (i) Das Bindungsproblem und die Ähnlichkeit von Anaphern und Präsuppositionen zeigen nicht, daß zweidimensionale Erklärungen falsch sind, sie zeigen nur, daß präsupponierte Information nicht ausschließlich Information über das Thema der Konversation beinhaltet. Getrennte Berechnung von Behauptung und Präsupposition können keine Bedeutungen ergeben, die Inhalte gemäß der Wahrheitsbedingungen repräsentieren. Diese Bedeutungen müssen abstraktere Gebilde sein, damit die Wahrheitsbedingungen der Behauptung von der Präsupposition abhängen können. (ii) Eine zweidimensionale Analyse der Präsupposition ist nicht nur möglich, sie ist auch wünschenswert, da sie die Trennung von semantischem Inhalt und Präsupposition ermöglicht. Anders als andere Ansätze kann eine zweidimensionale Analyse außerdem auf natürliche Weise Präsuppositionen quantifizierter Sätze erklären.

Ich argumentiere, daß eine quantifizierte Aussage wie *Jeder Deutsche liebt sein Buick* nur die Menge jener Deutschen betrifft, die einen Buick haben. Der Sprecher präsupponiert, daß diese Menge für die Konversation salient sei. Wenn er in einem bestimmten konversationellen Kontext nicht präsupponieren kann, daß solch eine saliente Menge Buick-besitzender Deutscher existiere, wird die Behauptung als unzutreffend gewertet, da keine Klarheit über die vom Satz ausgedrückte Proposition bestehen kann. Ich zeige, daß neuere eindimensionale Erklärungen von Präsuppositionen dieser Intuition nicht genügen, und demonstriere, wie sie auf zweidimensionalem Wege ohne die Entstehung von Bindungsproblemen erklärt werden kann.

Was Erlaubnissätze betrifft, werde ich erforschen, inwiefern man sie im Sinne der Mögliche-Welten-Semantik erklären kann. Ich werde argumentieren, daß für die Erklärung der Intuition für einen Satz des Erlaubens wie *Du darfst drei Äpfel nehmen.*, die ergibt, daß *höchstens* drei Äpfel genommen werden dürfen, wir die Dynamik des Erlaubten ernst nehmen müssen. Dieser Wandel dessen, was erlaubt ist, sollte dann teilweise im Sinne des zur Erklärung des Glaubenswandels entwickelten Thesenapparats, welcher in Kapitel 5 diskutiert wird, bestimmt werden. Ich werde aber vorschlagen, daß eine Erklärung von Erlaubnissätzen durch mögliche Welten insbesondere für konjunktive Erlaubnissätze problematisch ist. Eine Möglichkeit zur Lösung des Problems besteht in der Erklärung von Erlaubnissätzen unter ernsthafter Berücksichtigung von Tatsachen und Ereignissen.

Schließlich diskutiere ich kurz einen möglichen Weg zur Erklärung faktiver Verben im röckgriff auf Tatsachen und setze dies zur doppelt indizierenden Counterpart-theorie, die am Ende des ersten Kapitels entwickelt worden ist, in Beziehung.

Lebenslauf

Robert van Rooy
Eberhardstraße 41/43
70173 Stuttgart
Tel.: 0711- 233245

29.08.1966	Geboren in Drunen, Niederlande Vater: Harry van Rooy Mutter: Riet van Rooy, geb. Robben	
1972 - 1978	Grundschule, Haarsteeg	(Niederlande)
1978 - 1984	weiterführende Schule, Drunen	(Niederlande)
1984 - 1988	Landwirtschafliche Hochschule, s' Hertogenbosch Abschluß: Ingenieurdiplom (Ing.)	(Niederlande)
1988 - 1992	Studium der Philosophie 1988 - 1990 an der Katholischen Universität Tilburg 1990 - 1992 an der Katholischen Universität Nijmegen Abschluß (Nijmegen): Cum Laude	 (Niederlande) (Niederlande)
1991 - 1993	Studium der Sprachwissenschaft an der Katholischen Universität Tilburg Abschluß: Cum Laude	(Niederlande)
1993	Gaststudent am Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart mit einem Stipendium der V.S.B.-Bank aus die Niederlande	
1994 - 1997	Promotion im Graduiertenkolleg "Linguistische Grundlagen für die Sprachverarbeitung" am Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart Thema: "Attitudes and Changing Contexts"	
1997	Wissenschaftlicher Mitarbeiter der Forschergruppe "Logik in der Philosophie" an der Universität Tübingen Projekt: " Zur logischen Form von Glaubenszuschreibungen"	

Ich versichere, daß ich die vorliegende Dissertation mit dem Titel:

"Attitudes and Changing Contexts"

selbständig verfaßt und keine anderen Hilfsmittel als die angegebenen verwendet habe.

Stuttgart, den 9. Oktober 1997.

Bisher sind erschienen:

1 W. Frey: Syntaktische Bedingungen für die Interpretation – Über Bindung, implizite Argumente und Skopus (1990). Erschienen als Band 35 in der Reihe Studia Grammatica, Akademie-Verlag, Berlin.
2 S. Trissler, J. Pafel, M. Reis, J. Meibauer: Aspekte von W-Fragesätzen (1991).
3 R. Mayer: To Win and to Lose – Linguistic Aspects of Prospect Theory (1991).
4 T. Höhle: Koordinationsphänomene (1993).
5 J. Pafel: Zum relativen Quantorenskopus im Deutschen (1991).
6 M. Reis, I. Rosengren, J. Pafel: Weitere Aspekte von W-Fragesätzen (1991).
7 M. Reis, F.-J. d'Avis, J. Pafel, I. Rapp, S. Trissler, U. Lutz: W-Phrasen, W-Merkmale, Skopusberechnung (1992).
8 J. Wedekind: Unifikationsgrammatiken und ihre Logik (1991).
9 D. Kohl: Generierung aus unter- und überspezifizierten Merkmalsstrukturen in LFG (1991).
10 W. Kasper: Semantische Repräsentation und LFG (1992).
11 J. Geilfuß: Verb- und Verbphrasensyntax (1991).
12 A. Hestvik: Papers on Anaphors and Pronouns (1991).
13 N. Asher: Two Theories of Propositional Quantification / Abstract Entity Anaphora, Parallelism and Contrast (1991).
14 S. Berman, A. Hestvik: LF: A Critical Survey (1991).
15 A. Heilmann: Argumentstruktur (1991).
16 H. Kamp, U. Reyle: A Calculus for First Order Discourse Representation Structures (1991).
17 H. Haider: Fakultativ kohärente Infinitivkonstruktionen im Deutschen (1991).
18 S. Winkler: Subjektprominenz und sekundäre Prädikation (1991).
19 J. Delin: Aspects of Cleft Constructions in Discourse (1992).
20 C. Féry: Focus, Topic and Intonation in German (1992).
21 H. Kamp, A. Roßdeutscher: Remarks on Lexical Structure, DRS-Construction and Lexically Driven Inferences (1992).
22 H.-B. Drubig: On Topicalization and Inversion (1992).
23 H. Haider: Branching and Discharge (1992).
24 J. Delin: Towards a Taxonomy of Copular Constructions (1992).
25 M. Johnson, M. Kay: Parsing and Empty Nodes (1992).
26 M. Morreau: Conditionals in Philosophy and Artificial Intelligence (1992).
27 J. Geilfuß: Zur Kasuszuweisung beim AcI / Ist wollen ein Kontrollverb oder nicht? (1992).
28 F. Hamm: Ereignisstrukturen und Nominalisierungen (1992).
29 S. Berman, A. Hestvik (eds.): Proceedings of the Stuttgart Ellipsis Workshop. March 20–22, 1992 (1992).
30 P. Bosch, P. Gerstl (eds.): Discourse and Lexical Meaning. Proceedings of a Workshop of the DFG Sonderforschungsbereich 340, Stuttgart, November 30th - December 1st, 1992 (1992).
31 P. J. King, T. Götz: Eliminating the Feature Introduction Condition by Modifying Type Inference (1993).
32 S. Lorenz: Temporales Schließen unter Standardannahmen bei der Verarbeitung natürlicher Sprache (1993).
33 A. Roßdeutscher: Using Lexical Information to Construct Discourse Representation Structures: A Case Study in Anaphora Resolution (1993).
34 F.-J. d'Avis, S. Beck, U. Lutz, J. Pafel, S. Trissler: Extraktion im Deutschen I (1993).
35 G. Minnen, D. Gerdemann: Direct Automated Inversion of Logic Grammars (1993).
36 D.M. Gabbay: Fibred Semantics and the Weaving of Logics 1 (1993).
37 A. Merin: Lexicality, Indexicality, and Compositional Contrarieties: A Case Study on Aristotle (1993).
38 A. Merin: Algebra of Elementary Social Acts (1993).
39 S. Beck: Interventionseffekte für LF-Bewegung (1993).
40 T. Götz: Unique Normal Forms and Subsumption For Typed Feature Structures (1994).
41 H. Haider: Detached Clauses – The Later The Deeper (1994).
42 A. Heilmann: Katalog der Datenbanken (1994).
43 A. Frank: Verb Second by Lexical Rule or by Underspecification (1994).
44 J. Möck: Extraposition aus der NP im Englischen (1994).
45 G. Minnen: Predictive Left-to-Right Parsing of a Restricted Variant of TAG(LD/LP) (1994).
46 I. Kohlhof: Diskurskohärenz und Akzentuierung bei adnominaler Quantifikation im Deutschen (1994).
47 J. Bayer: Barriers for German (1994).
48 J. Dörre: Feature-Logik und Semiunifikation (1994).
49 L. Chen, C. Féry: Thetische und kategorische Sätze im Mandarin (1994).
50 J. Delin: On Processing Clefts (1994).
51 H.-B. Drubig: Island Constraints and the Syntactic Nature of Focus and Association with Focus (1994).
52 U. Kleinhenz: Focus and Phrasing in German (1994).
53 P. Gerstl: Die Berechnung von Wortbedeutung in Sprachverarbeitungsprozessen. Possessivkonstruktionen als Vermittler konzeptueller Information (1994).
54 E. König: A Study in Grammar Design (1994).
55 J. Meier: Zur Syntax des Verbalkomplexes im Deutschen (1994).
56 B. Mergel: Wortstellungsalternation im deutschen Mittelfeld (1994).
57 M.-W. Choi: Kasustheorie und Mehrfachnominativkonstruktionen im Koreanischen (1994).
58 E. Hinrichs, D. Meurers, T. Nakazawa: Partial-VP and Split-NP Topicalization in German – An HPSG Analysis and its Implementation (1994).
59 P. J. King: An expanded logical formalism for head-driven phrase structure grammar (1994).
60 S. Kepser: A Satisfiability Algorithm for a Typed Feature Logic (1994).
61 S. Berman: Wh-Clauses and Quantificational Variability: Two Analyses (1994).
62 P.A. Schindler: Dritter Ton Sandhi im Mandarin und Prosodische Struktur (1994).
63 C. Fortmann: Zur w-Syntax im Deutschen (1994).
64 S. Winkler: Secondary Predication in English: A Syntactic and Focus-Theoretical Approach (1994).
65 S. Berman, A. Hestvik: Principle B, DRT and Plural Pronouns (1994).
66 G. Müller: A Constraint on Remnant Movement (1994).
67 F.-J. d'Avis: Zu selbständigen und-eingeleiteten Verbletzt-Sätzen im Deutschen (1995).
68 F. Morawietz: Formalization and Parsing of Typed Unification-Based ID/LP Grammars (1995).

69 I. Kohlhof, S. Winkler, H.-B. Drubig (eds.): Proceedings of the Göttingen Focus Workshop, 17. DGfS, March 1-3, 1995 (1995).
70 H. Haider: Studies on Phrase Structure and Economy (1995).
71 M. Bierwisch, P. Bosch (eds.): Semantic and Conceptual Knowledge. Proceedings from a joint workshop of the Arbeitsgruppe Strukturelle Grammatik and the Institut für Logik und Linguistik, April 21-23, Berlin 1994 (1995).
72 J. Rogers: Capturing Linguistic Theories Model-Theoretically (1996).
73 S. Beck, S.-S. Kim: On Wh- and Operator Scope in Korean (1996).
74 P.A. Schindler: The Syntax of Negative Polarity (1996).
75 A. Merin: Formal Semantic Theory and Diachronic Data: A Case Study in Grammaticalization (1996).
76 U. Lutz, G. Müller (eds.): Papers on Wh-Scope Marking. Papers from the workshop The Syntax and Semantics of Wh-Scope Marking, December 1-2, Tübingen 1995 (1996).
77 R. Eckardt: Intonation and Predication: An investigation in the nature of judgement structure (1996).
78 A. Merin: Neg-Raising, Elementary Social Acts, and the Austinian Theory of Meaning (1996).
79 T.-Cornell: A Minimalist Grammar for the Copy Language (1996).
80 C. Grefe, M. Kracht: Adjunction Structures and Syntactic Domains (1996).
81 L. Kallmeyer: Underspecification in Tree Description Grammars (1996).
82 H. Volger: Principle Languages and Principle Based Parsing (1997).
83 T. Cornell: Representational Minimalism (1997).
84 J. Rogers: The Descriptive Complexity of Generalized Local Sets (1997).
85 F Morawietz, T.L. Cornell: On the recognizability of relations over a tree definable in a monadic second order tree description language (1997).
86 F. Morawietz: Monadic Second Order Logic, Tree Automata and Constraint Logic Programming (1997).
87 T. Fernando: Papers on Dynamic Semantics and Related Topics (1997).
88 A. Merin, C. Bartels: Decision-Theoretic Semantics for Intonation (1997).
89 I. Reich: Wer will wann wieviel wissen? Eine Untersuchung verschiedener Frage-Antwort-Bedingungen im Deutschen (1997).
90 F.-J. d'Avis, U. Lutz: Zur Satzstruktur im Deutschen (1997).
91 A. Frank: Context Dependence in Modal Constructions (1997).
92 A. Palm: The Expressivity of Tree Languages (1997).
94 L. Chen, W. Schaffar: Ja/Nein-Fragen im Mandarin, im Xiang und im Thailändischen (1997).
95 E. Hinrichs, D. Meurers, F. Richter, M. Sailer, H. Winhart: Ein HPSG-Fragment des Deutschen. Teil 1: Theorie (1997).
96 E. König: A Matrix Proof Method for Underspecified Discourse Representation Structures (1997).
97 W. Sternefeld: The Semantics of Reconstruction and Connectivity (1997).
98 M. Brody: Projection and Phrase Structure (1997).
100 A. Merin: Information Relevance and Social Decisionmaking: Some Principles and Results of Decision-Theoretic Semantics (1997).
101 A. Merin: If all our arguments had to be conclusive, there would be few of them. (1997)
102 A. Merin: On the Language and Cognitive Dynamics of Proof (1997).
103 D. Abusch: Generalizing Tense Semantics for Future Contexts (1997).
104 A. Wöllenstein-Leisten, A. Heilmann: The Syntax-Semantics Interface Conditions for Infinitival Complementation (1997).
105 C. Meier, I. Kohlhof: Presupposition and Information Structure (1997).
106 E. Goebbel: On the Double Object Construction (1997).
107 P. Gallmann: Zu Morphosyntax und Lexik der w-Wörter (1997).
108 D. Abusch, M. Rooth: Epistemic NP Modifiers (1997).
109 H.-B. Drubig: Some Cross-Categorial Generalizations of Focus Structure (1997).
110 H.-P. Kolb: GB Blues. Two Essays on Procedures and Structures in Generative Syntax (1997).
111 U. Mönnich: Zur bereichstheoretischen Semantik von Grammatikformalismen (1997).
112 H. van Hoof: On Split Topicalization and Ellipsis (1997).
113 H.-P. Kolb: Macros for Minimalism. Towards Weak Descriptions of Strong Structures (1997).
114 U. Mönnich: On Cloning Context-Freeness (1997).
115 H.-B. Drubig: Fokuskonstruktionen (1997).
116 A. Paslawska: Transpararente Morphologie und Semantik eines deutschen Negationsaffixes (1997).
118 M. Hiller: Baumzulassung ist kontextfrei (1997).
119 R. Meyer: Extraktionsbeschränkungen im Deutschen und Russischen: Deklarativsatzkomplemente und Nominalphrasen (1997).
120 K. Eberle: Flat underspecified representation and its meaning for a fragment of German (1997).
124 B.-R. Ryu: Argumentstruktur und Linking im Constraint-basierten Lexikon: Ein Zwei-Stufen-Modell für eine HPSG-Analyse von Ergativität und Passivierung im Deutschen (1997).
125 R. van Rooy: Attitudes and Changing Contexts (1997).