Belief in Context:
Towards a unified semantics of de re and de se attitude reports

een wetenschappelijke proeve op het gebied
van de Filosofie

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de Rector Magnificus
prof. dr. C.W.P.M. Blom
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
donderdag 30 november 2006
om 10.30 uur precies
door

Emar Maier

geboren op 30 mei 1977 te Amsterdam
Promotor:
   Prof. dr. R.A. van der Sandt
Copromotor:
   Dr. B. Geurts

Manuscriptcommissie:
   Prof. dr. J.A.G. Groenendijk  Universiteit van Amsterdam
   Prof. dr. M.V.P. Slors
   Prof. dr. T.E. Zimmermann  J.W. Goethe Universität Frankfurt

© 2006 Emar Maier
ISBN 90-81146-1-1
Typeset with LATEX 2ε
Cover design by Jennifer Spenader
Printed by Print Partners Ipskamp, Enschede

The work in this thesis has been partially funded by the Netherlands Organization for Scientific Research (NWO), grant 360.80.013.
Acknowledgments

I would like to thank everybody who helped me in any way over the past five years, or whom I just hung out with. In order to make some distinction among these, let me mention a few names and leave the rest implicit.

First of all I want to thank my advisors Rob van der Sandt and Bart Geurts for all their help. They’ve been responsible for introducing me to the field of semantics, and, just as importantly, to some of the leading figures in that field. Among these, the members of my thesis committee, Jeroen Groenendijk, Ede Zimmermann and Marc Slors, who I thank for agreeing to be in that committee and for reading the, rather lengthy, manuscript in such a short time. But that manuscript wouldn’t have been there if it wasn’t for the discussions I’ve had with a number of semanticists, of whom I’d like to mention three in particular: Philippe Schlenker, who introduced me to monsters, and to its key literature and researchers, and who was always very helpful and supportive; Ede Zimmermann, who, among other things, invited me to Frankfurt and helped me fix a major hole in my analysis; and Henk Zeevat, who often had interesting comments and ideas and liked to discuss them. Still, I have received even more help from my close friends (and colleagues) Corien Bary and Peter de Swart. Additional thanks go to Janneke Huitink, Hanneke van der Grinten and the rest of our TFL department, and to Helen de Hoop, Nick Asher, and everybody at the ILLC in Amsterdam. I also thank the anonymous referees and audiences of various conferences for their many insightful questions, comments and criticism.

Finally, I’d like to thank my family and friends (in so far as they are not already included in any of the other categories). A huge thanks and 000s go to my girlfriend Jennifer Spenader. This book is dedicated to K’Kruhk, the favorite Jedi of my son Ward van Oosterhout. May The Force be with you.
Contents

Acknowledgments iii

Introduction xi

1 Attitudes 1

1.1 Introduction: semantics and the attitudes 1

1.1.1 Propositions, possible worlds, contexts, and attitudes 1

1.1.1.1 Syntax, semantics, and pragmatics 2

1.1.1.2 Logical form 3

1.1.1.3 Compositionality 5

1.1.2 Propositional attitudes 7

1.1.2.1 Doxastic logic 9

1.1.2.2 Possible worlds and representation 9

1.1.2.3 Logical omniscience and Frege’s puzzle 11

1.1.3 Modes of attitude: de dicto, de re, de se 15

1.1.3.1 De dicto 15

1.1.3.2 De re 16

1.1.3.3 De se 17

1.2 Modes of attitude 19

1.2.1 De dicto attitudes: the disquotational principle 20

1.2.2 De re attitudes: singular vs. descriptive content 22

1.2.2.1 Singular propositions 23

1.2.2.2 Substitution 25

1.2.2.3 Belief under a description and the shortest spy problem 28

1.2.2.4 Acquaintance relations 30

1.2.2.5 Kaplan (1969): reducing de re to de dicto 32

1.2.3 De se attitudes: indexical belief as self-ascription of properties 35

1.2.3.1 The essential indexical 36

1.2.3.2 De se vs. de dicto: properties vs. propositions 41
## Contents

1.2.3.3 *De se* vs. *de re*: more terminological remarks 47
1.2.4 Unifying *de re* and *de se* 49  
  1.2.4.1 Relational attitudes 49  
  1.2.4.2 Kaplan: character vs. content 55  
  1.2.4.3 Diagonals: from characters to properties 59  
  1.2.4.4 Character theory vs. relational attitudes 62

1.A Modal predicate logic 64  
  1.A.1 Syntax 64  
  1.A.2 Models 65  
  1.A.3 Free variables and assignments 65  
  1.A.4 Semantics 65  
  1.A.5 Extra modalities: belief and tense 67

1.B Kaplan’s 2-dimensional semantics 67  
  1.B.1 Syntax 67  
  1.B.2 2D models 68  
  1.B.3 Semantics 68  
  1.B.4 Adding ‘says’ and ‘believes’ 69

2 Attitude reports 71

2.1 *De dicto*, *de re*, and *de se* reports? 71  
2.2 *De dicto* and *de re* reports 75  
  2.2.1 Scope ambiguity 75  
  2.2.2 Kaplan (1969) revisited: vivid names and *de re* as *de dicto* 77

2.3 *De se* and *de re* 79  
  2.3.1 The semantics of relational attitude 79  
  2.3.2 The pragmatics of acquaintance 85  
    2.3.2.1 Contextualized acquaintance 86  
  2.3.3 Compositionality? 89  
    2.3.3.1 Structured propositions in logical forms 90

2.4 Anti-reductionism and pure *de se* separatism 110  
  2.4.1 Pure *de se* reports 112  
    2.4.1.1 Indirect reflexives and *he* 112  
    2.4.1.2 Logophors 115  
    2.4.1.3 Chierchia’s PRO and the ambiguity thesis 118  
    2.4.1.4 Abstract anaphora: Chierchia vs. the hybrids 123
3.2.2.1 Two-stage interpretation .................................. 202
3.2.2.2 Binding presuppositions .................................. 205
3.2.2.3 Accommodation ............................................. 212
3.2.3 Presuppositions in attitude reports ......................... 218
   3.2.3.1 Presupposition as scope in de dicto/de re reports ..... 220
   3.2.3.2 De re as de dicto in DRT .............................. 226
   3.2.3.3 From vivid names to acquaintance relations .......... 232
3.3 Layered DRT ...................................................... 237
   3.3.1 Basic LDRT .................................................. 240
      3.3.1.1 Syntax ................................................. 240
      3.3.1.2 Semantics ............................................. 241
   3.3.2 Direct reference vs. descriptivism in LDRT .............. 243
      3.3.2.1 Against descriptivism ................................. 243
      3.3.2.2 Descriptivism revisited: proper names as presuppositions ... 247
      3.3.2.3 Real rigidity in DRT: external anchors .......... 250
      3.3.2.4 A two-dimensional semantics of the k layer ....... 252
   3.3.3 Presupposition and the two-stage architecture in LDRT 255
      3.3.3.1 A layer for presuppositions? ...................... 256
      3.3.3.2 Cross-layer binding and non rigid proper names .... 260
   3.3.4 Belief in LDRT .............................................. 270
      3.3.4.1 Syntax ................................................. 270
      3.3.4.2 Semantics ............................................. 271
3.4 Belief reports in LDRT .......................................... 273
   3.4.1 Recap: de dicto, de re, and de se reports in LDRT ...... 273
      3.4.1.1 Known issues ......................................... 277
   3.4.2 Acquaintance Resolution .................................. 280
      3.4.2.1 Higher-order unification ............................. 280
      3.4.2.2 The resolution of acquaintance ..................... 282
      3.4.2.3 Applications: Ortcutt and mistaken identity ....... 285
   3.4.3 Quantifiers .................................................. 292
      3.4.3.1 Universal quantification .............................. 293
      3.4.3.2 Only .................................................. 302
   3.4.4 Pure de se reports and shifted indexicals ............... 308
      3.4.4.1 Acquaintance with concepts? ......................... 310
      3.4.4.2 Introspection: a rationality criterion ............. 316
      3.4.4.3 Conservative? ....................................... 321
      3.4.4.4 PRO and LOG again .................................. 324
## Contents

4 Conclusions and discussion  329  
  4.1 Conclusions  329  
  4.2 Future research  332  

Glossary of names  335  

Bibliography  353  

Summary  355  

Geloof in context: samenvatting  365  

Curriculum Vitae  369
Introduction

This thesis deals with the phenomenon of attitude reporting. More specifically, it provides a unified semantics of *de re* and *de se* belief reports. After arguing that *de se* belief is best thought of as a special case of *de re* belief, I examine whether we can extend this unification to the realm of belief reports. I show how, despite very promising first steps, previous attempts in this direction ultimately fail with respect to some relatively recent linguistic data involving quantified and infinitival reports, logophoric constructions, and monstrously shifted indexicals. Formalizing my idea of a contextual resolution of acquaintance relations in a dynamic framework, I arrive at an alternative analysis that handles all these data.

What follows is an outline of the organization of the thesis. Very briefly, the first chapter deals with attitudes, focusing on the *de re* and *de se* modes of the attitude of belief. The second chapter deals with the way natural languages can be used to report someone’s attitudes and describes the classical semantic analyses of these linguistic constructions called attitude reports. Finally, the third chapter is devoted to my own proposal for a report semantics, one that tries to unify *de re* and *de se* by combining some of the classical ideas with a truly dynamic account of context dependence.

Chapter 1 starts off by introducing such foundational concepts as semantics and pragmatics, logical form, and compositionality, right up to the analysis of propositional attitudes in possible worlds semantics and its main problems. From there I define the modes of attitude, *de dicto*, *de re*, and *de se*. Then I review the logical apparatus that has been used to characterize these classes formally. *De dicto* is the basic mode, characterized by the so-called disquotational principle: if you are disposed to assent to a sentence $\varphi$, then you believe the proposition expressed by that $\varphi$. Trying to pin down *de re* brings us from acquaintance, singular propositions and rigidity to vivid names and belief under a description. The perspectival, non-propositional distinctions underlying *de se* belief are captured by going from propositions to properties and from worlds to contexts. I finish this introduction to attitude semantics by comparing two well-known unified analyses of *de re* and *de
Chapter 2 is about attitude reports, i.e. (English) sentences typically of the form *x believes that such and such* that are meant to convey that someone believes something. The first observation is that we cannot always see from the linguistic surface form of the report what mode of attitude it reports, i.e. many reports are ambiguous (or underspecified) with respect to the mode of the reported attitude. Then I introduce some classic accounts of report semantics, focusing on those that deal with the *de re/de se* interface. First, I consider a version of the relational analysis adapted to reports, and discuss two defects and possible enhancements (context dependent acquaintance relations and structured propositions). When we do the same with the two-dimensional Kaplanian alternative we discover that the Kaplan-style report semantics is fatally flawed. Next, we widen our data set by adding reports with logophoric pronouns, infinitival complements, quantified reports and monstrously shifted indexicals in various languages. Most of these data have been used to argue against the total unification of *de re* and *de se* (relational or otherwise). These data have also given rise to a host of alternative proposals that tease *de re* and *de se* reports apart. Comparing them and the relational analysis, I conclude that all have theoretical and/or empirical disadvantages, thus clearing the way for a new proposal.

Chapter 3 presents my own account, based on the relational one, but set in a dynamic framework to properly implement the idea of a contextually determined acquaintance relation. I start by developing and defending the general dynamic framework that seems most suitable. I start with a standard DRT, adding quantification, intensionality and presupposition (“as anaphora”). To capture the *de re/de se* distinction we also need a semantics of directly referential terms, *I* in particular. For this purpose I use a novel DRT extension: Layered DRT, a framework for representing different types of informational content on separate yet interconnected layers within a single DRS. Combining this layering and its two-dimensional semantics with presupposition theory leads to a very promising new semantics of indexicals and proper names, combining insights from Russellian descriptivism with Kripke-Kaplanian direct reference. Then we return to the original issue: to devise a sound, unified semantics of *de re/de se* reports, based on relational attitudes, where the acquaintance relation is contextually determined. My proposal is called Acquaintance Resolution. The key ingredient is an underspecified acquaintance relation that gets bound by a salient relation from the context. This process of binding is formalized as a generalization of regular presupposition binding, using higher-order unification. Two subsequent extensions are
added to this simple analysis in view of the previous chapter’s problematic data: quantified reports, logophors, infinitives and shifted indexicals. The end results is a unified semantics of *de re* and *de se* belief reports, formalized in a dynamic framework, and able to handle the wide variety of otherwise problematic data that has accumulated in the literature on the subject.
Chapter 1

Attitudes

1.1 Introduction: semantics and the attitudes

This section introduces some of the basics of the study of attitudes, immediately narrowing it down to the attitude of belief. In 1.1.1 I review some terminology and background for the kind of semantics we’ll be dealing with. Then, in 1.1.2, we turn to the subject of attitudes and discuss the traditional analysis in terms of modal logic and possible worlds semantics. In 1.1.3, the distinctions between the three classical modes of attitude, de dicto, de re, and de se, are discussed from a pre-theoretical point of view.

1.1.1 Propositions, possible worlds, contexts, and attitudes

Human beings can have various attitudes towards all kinds of information. Philosophers traditionally take propositions as the unit of information, especially of information conveyable by means of language. More specifically, the informational content of an utterance is called the proposition expressed by that utterance. On a standard correspondence theory of truth we define true as in correspondence with the facts (states of affairs in the actual world). The notion of a proposition expressed by an utterance can now be further explicated as the set of all possible states of affairs that would make it true, i.e. the proposition expressed by an utterance is the set of worlds in which that utterance is true. The intersection of linguistics, logic and philosophy concerned with propositions is called (truth-conditional) semantics. The current section is meant to introduce the key notions of the type of semantic study exemplified by the rest of this work.
1.1.1.1 Syntax, semantics, and pragmatics

On the linguistic dimension, semantics is often situated between two other disciplines: syntax and pragmatics. Roughly speaking, syntax deals with (well-formedness of) sentences, semantics deals with (truth conditions of) utterances and pragmatics is concerned with what people do with language in the real world. Syntax is interested in sentences as formal, structured objects and tries to characterize the set of well-formed (grammatical, interpretable) sentences of a given natural language. Pragmatics, among other things, is interested in in the way language depends on the external context in which an utterance takes place, and in speech acts other than assertion that arguably are not meant merely to convey a certain amount of information (e.g. questions and promises). Bluntly put, syntax asks whether a sentence is well-formed, semantics whether an utterance (i.e. a sentence a uttered in a context) is true, pragmatics whether it’s felicitous/acceptable in a given situation or discourse.

To be a bit more precise then, the current work may be provisionally situated within the syntax-semantics interface, studying how to determine systematically from the surface structure of a particular class of sentences (viz. attitude reports) what propositions they express. Many, but certainly not all, authors studying attitude reports would indeed consider themselves to work at that interface. My own proposal, laid out in chapter 3, and some of its predecessors to be encountered in chapter 2, however, essentially involve presuppositions and other varieties of context dependence, phenomena often classified as pragmatic or pragmasemantic.\(^1\) In so far as presupposition resolution is governed by pragmatic felicity constraints, the label pragmatics is evidently justified. However, we refrain from trivializing semantics by the once popular definition of pragmatics as ‘everything involving context-dependence’. It is after all utterances (sentences uttered in context) that carry truth conditions and express propositions, not mere sentences. For, when is “She never did.” taken in isolation true? Indeed, most if not all sentences are context-dependent in the sense that they require some features of the utterance situation to make them express any, let alone a single identifiable meaning. A case in point are anaphoric pronouns (in the broad sense of the word, covering both inter- and intrasentential anaphora) that require a previous discourse context to give them an antecedent to link up to. Other typical examples involving context dependence are the reference determination of demonstratives and other forms of indexicality (deixis), and the disambiguation of lexical items.

\(^1\)In fact, one of the contributions of this work is to shift the focus of the study of attitude reports from syntax to pragmatics.
1.1 Introduction: semantics and the attitudes

The above considerations concerning context dependence, be it termed semantic or pragmatic, also show the need for a very broad notion of context. Contexts will need to span several dimensions, among which the (information conveyed by) preceding discourse, world knowledge, encyclopedic knowledge, and also the real physical context surrounding the utterance situation in space-time. Earlier theories often focus only on one aspect of context. Kaplan’s (1989) famous logic of demonstratives for instance construes contexts as centered possible worlds, modeling external utterance situations, whereas a more recent trend called dynamic semantics takes contexts to be the information conveyed by previous discourse and moreover shifts the primary focus of semantics from the proposition expressed by a sentence to the effect it has on such a context. This chapter and the next will be concerned entirely with so-called static, i.e. truth-conditional, semantics.

1.1.1.2 Logical form

This thesis is about interpreting natural language sentences that report someone’s beliefs. As a semanticist, the data I am concerned with consist of truth value judgments of sentences in (hypothetical) contexts. The main theoretical questions then are, first, how to represent the truth conditions of such sentences, and second, how to get to the truth conditions of these sentences in a systematic way.

As for the first question, truth conditions must be represented in some formal language which has an exact model-theoretic truth definition. Examples of formal systems that have been used for this task are first-order predicate logic (e.g. Russell, Quine), or something more powerful like $\lambda$-categorial grammar interpreted in an intensional theory of types (Montague). Following Russell, these formal representations of truth conditions have been termed logical forms. In most systems the logical form of a sentence differs from its realization in a natural language speech act, the so-called surface structure. In fact, a single surface structure often admits various logical forms, viz. in the case of a structurally ambiguous sentence. A case in point is (1), which also introduces my notation $\triangleright$ to mark a sentence’s logical form (usually

\footnote{Be careful not to confuse this Russellian notion of logical form with the Chomskyan Logical Form (LF). Terminology with respect to ‘lf vs. LF” is generally a mess in the literature. I will talk about logical formulas representing truth conditions as logical forms and reserve the capitalized expression for a distinct level of syntactic representation. The interested reader can find more on this distinction in most semantics textbooks, e.g. Chierchia & McConnell-Ginet (1990:144-5) or Gamut (1991:214-5). Note, however, that in some semantic frameworks no such distinction is made and objects called LF fulfill both roles, or it is claimed that there is only one level of syntactic structure, viz. phrase structure trees, and that is isomorphic to logical form (Montague).}
written in some kind of predicate logical formalism):

(1) Every student speaks a certain language

a. $\forall x [\text{student}(x) \rightarrow \exists y [\text{language}(y) \land \text{speak}(x, y)]]$

b. $\exists y [\text{language}(y) \land \forall x [\text{student}(x) \rightarrow \text{speak}(x, y)]]$

The two predicate logical formulas in (1) represent the distinct truth conditions for the two readings of the allegedly ambiguous English sentence. The first represents the reading in which for every student there is a language she speaks, where this language may differ from student to student. The second formula is only true in case every student speaks the same language. This reading can be made more salient by adding “But I don’t know which one”. This particular type of structural ambiguity is called scope ambiguity: the different readings correspond to the two ways the quantifiers of the logical form can take scope. Note that the converse is also true: a single logical form can be expressed in various ways:

(2) $\triangleright \text{kiss}(c, l)$

a. Colby kisses Ligia

b. Ligia was kissed by Colby

From the examples above it will have become clear that despite some similarities, natural language sentences and their logical representations are quite far apart. Frege, the inventor of predicate logic with its variables and quantifiers, was the first to note that superficially similar expressions of natural language like Colby and everybody, or exists and walks have completely different functions from a logical point of view. Syntactically, both Colby and everybody are noun phrases (NP), while exists and walks are both verb phrases (VP), meaning that they have the same distribution (replacing one with the other within an English sentence will preserve well-formedness). But logically, Colby is a simple referring expression (c), everybody involves quantification ($\forall x [\text{person}(x) \rightarrow ]$), as does exists ($\exists x [ ]$), and finally walks is a predicate (walk).

The conclusion Frege draws from the observed divergence between logical and natural language was quite moderate; he compares the distinction to that between the microscope and the eye, the first very—but exclusively—useful for scientific explorations, the other for everyday life (Frege 1879:xi). Russell (1905) and Wittgenstein (1921), and in their wake the Vienna Circle, take this a step further, to the point where, under the header Misleading Form Thesis, they blame natural language for being full of ambiguities and other imperfections. Their program was to replace natural language with
something based on predicate logic, especially in the areas of scientific and philosophical discourse.

To sum up, there is a prima facie distinction between a sentence as used by speakers of a natural language, and the logical form encoding its truth conditions in a formal language. Now that we have established this discrepancy, let’s turn to the second question, which concerns the relation between surface and logical form: How can we systematically derive logical forms from surface structures?

1.1.1.3 Compositionality

Notice the previous sentence’s use of *systematically*, presupposing that there are systematic connections between surface and logical forms. Thus, in contrast to the Misleading Form Thesis, we’ll assume that there is a correspondence between the logical operators, predicates, quantifiers etc. that make up a logical form and the grammatical categories that make up the syntactic structure. In fact, I take it that one of the main desiderata for syntactic theory should be that its output structures can be systematically mapped on a logical form representing the sentence’s truth conditions. Other syntax internal constraints on syntactic theory, involving such notions as grammaticality and constituent structure, will not concern us in this work.

The notion of systematicity alluded to above is often explicated as *compositionality*, i.e. the principle that the meaning of a compound expression is determined solely by the meanings of its parts and the way these are put together. This notion of compositionality derives from Frege who argues as follows:

It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a terrestrial being for the first time can be put into a form of words which will be understood by someone to whom the thought is entirely new. This would be impossible, were we not able to distinguish parts in the thoughts corresponding to the parts of a sentence, so that the structure of the sentence serves as the image of the structure of the thoughts.  

[Frege 1923:1]

This argument is often referred to as the argument from productivity: on the basis of a finite lexicon we can produce and understand infinitely many novel sentences. In a compositional language system it indeed suffices to learn some basic words plus the ways in which to compound them and we can produce an infinite number of sentences. If, in addition, we also know the meanings of the basic words, plus the way the meaning of a compound
depends on the meanings of its parts, we’re done. A slightly different way
to characterize this type of compositionality is to say there’s a rule-by-rule
correspondence between syntax and semantics: each grammar rule that tells
us how to combine a number of well-formed strings into a new well-formed
string, is accompanied by a semantic rule determining the interpretation of
the new string from the interpretations of the parts. We can illustrate the
principle of compositionality with the language of the propositional calculus
and its interpretation. In the syntax we have a rule that says you can combine
two formulas into a new one by concatenating them with a ‘wedge’ sign in
between:

\[(3) \text{ If } \varphi \text{ and } \psi \text{ are well-formed formulas, then } \varphi \land \psi \text{ is a well-formed formula}\]

This rule corresponds to the following semantic rule, wherein \([\varphi]\) denotes the
interpretation, i.e. the truth-value, of \(\varphi\) relative to some fixed model:

\[(4) \quad [\varphi \land \psi] = [\varphi] \cdot [\psi]\]

This rule pair is all there is to the formalization of the connective \(\land\) in
propositional calculus.

All interpreted formal languages we will encounter have their syntax and
semantics set up in this way. It is still heavily debated however whether it’s
necessary and desirable to demand that descriptions of natural language be
compositional in the same way. Montague was the first to express his con-
viction that there is no real difference between natural and formal language
in this respect, and that therefore the semanticist’s endeavor consists in giv-
ing a compositional syntax/semantics for (significant fragments of) natural
language. His formal descriptions cover an impressive fragment of English,
using a fairly straightforward (categorial) grammar coupled with an inten-
sional, typed semantic framework, a generalization of first-order predicate
logic making heavy use of lambda abstraction. This type of semantics with
its emphasis on compositionality has been very influential to this day and it’s
clear that a good deal of semantics at least can be done in a compositional
manner. For this reason I take compositionality to be one major (theory
internal) desideratum for semantics.

That said, I take it as equally well-established that there are aspects of
natural language interpretation that are decidedly non-compositional. Un-
controversial examples of this are dictionary meanings of words and stock
phrases, but more relevantly, certain context-dependent and pragmatic as-
pects of meaning. Presupposition is a prime example of a non-compositional
aspect of meaning. Consider (5) a sentence with a definite description, ar-
guably triggering the presupposition that there is a unique donkey owned by Ligia:

(5) Ligia beats her donkey

Assuming a very direct form of compositionality we would expect that the interpretation of (5) depends only on the interpretations of its constituents, say Ligia, beats, her, and donkey. The predicate logical representation traditionally assigned to such sentences in isolation, may well be derived compositionally, as demonstrated by Montague:

▶ ∃x[donkey(x) ∧ own(1, x) ∧ ∀y[donkey(y) ∧ own(1, y) → y = x] ∧ beat(1, x)]

But, it turns out that the linguistic context may play a crucial role as well, as brought out by the example in (6), which contains (5) as a constituent while having a logical form totally disjoint from (5):

(6) [I’m not sure if Ligia currently has any pets, but the fact is:] If she owns a donkey, Ligia beats her donkey

▶ ∀x[(donkey(x) ∧ own(1, x)) → beat(1, x)]

Note that from a purely technical point of view it may not be impossible to give a compositional account of these data, the point is that it would be a rather complicated and ad hoc account. More specifically, it would have less redeeming value for the study of natural language than some alternative analyses that violate strict rule-by-rule correspondence between surface structure and logical form while adhering to some kind of systematicity in deriving lf. These alternatives often resort to movements and other transformations in the syntactic module, or to non-meaning-preserving operations on the semantic end. I tentatively conclude that compositionality is an important consideration, though I’d immediately give up the strict Montagovian implementation of it if a phenomenon is better described in a non-compositional yet systematic fashion. Chapter 2 contains some in depth discussion of these compositionality issues in a concrete example featuring belief reports.

1.1.2 Propositional attitudes

With these musings on general semantic methodology in place, let us return to the key observation introducing the object of this study—that we can have attitudes towards propositions and that propositions can be understood as sets of possible worlds. For example, take an utterance of (7).
Chapter 1. Attitudes

(7) Ligia is a philosopher

This is a meaningful statement and as such it expresses a proposition, viz. the proposition that Ligia is a philosopher. This proposition can be identified with the set of possible worlds in which Ligia is a philosopher. In this case that proposition, call it \( p \), is a proper, non-empty subset of the set of all possible worlds \( W \):

\[
\begin{align*}
\text{(8) } \quad p & \subseteq W \\
& = \text{ the proposition expressed by (7)} \\
& = \text{ the proposition that Ligia is a philosopher} \\
& = \{ w \in W \mid \text{Ligia is a philosopher in } w \} \\
& = \llbracket \text{Ligia is a philosopher} \rrbracket
\end{align*}
\]

I can bear an attitude of belief to proposition \( p \), but I can also doubt or even fear it. This work is almost exclusively concerned with the attitude of belief and reports thereof, so let’s focus on that from now on.

The kind of semantics we envisage concerns itself with propositions as the basic units of interpretation of formulas, rather than just truth-values, and relativizes truth to truth at a possible world. In other words, we’ll be concerned solely with intensional systems, starting with modal predicate logic below.

But before we go on, a quick note to clarify some notational conventions used throughout. Propositions and sentences/utterances are distinguished typographically: I use quotation marking (italics, single or double quotation marks, or bracketed example numbers) for mentioning sentences and utterances; and I use that-clauses to mention propositions, as in (7) expresses the proposition that Ligia is a philosopher. If I need to abstract away from any particular sentence and use variables over expressions, I use Greek letters, like \( \sigma \) for natural language sentences, \( \varphi \) for formulas (never for propositions), and \( \alpha \) for variables. I use a typewriter font for expressions of formally defined object languages (\( \text{beat}(x, 1) \)), and, finally, italics and calligraphic letters to refer to model-theoretic entities like (sets of) individuals \( a \in D \), possible worlds \( w \in W \), and propositions \( p \subseteq W \). This typography parallels the logician’s strict division between (model-theoretically) ‘real’ and mere formal objects.

\[\text{\footnotesize\textsuperscript{3}It is important not to confuse the sentence (7), “Ligia is a philosopher”, with the proposition that its utterance expresses, viz. that Ligia is a philosopher, } p.\]
1.1 Introduction: semantics and the attitudes

1.1.2.1 Doxastic logic

The notion of believing a proposition can be fleshed out in modal logical terms by taking as a primitive model-theoretic notion a relation, $R$, of doxastic accessibility between possible worlds: $wRv$, pronounced “$v$ is doxastically accessible from $w$”, meaning that in $w$ the subject cannot distinguish $v$ from (her conception of) the real world.

(9) $a$ believes proposition $p$ at $w$ iff for all $v$ with $wRv$, $v \in p$

For example, Colby believes a proposition iff at all worlds doxastically accessible to him, that proposition is true.

A slightly different setup is one where we equate a person’s belief state with a set of possible worlds, her so-called doxastic alternatives, which we then take as the primitive notion, i.e. given by the model. If we denote a subject’s belief set at world $w$ as $Bel(w)$, we can define belief as:

(10) $a$ believes proposition $p$ at $w$ iff $Bel(w) \subseteq p$

Instead of saying that Colby believes that Ligia is a philosopher iff in every accessible world Ligia has the property of being a philosopher, we now define it by requiring that in every world of his belief set Ligia is a philosopher. Of course both definitions of belief, in terms of doxastic accessibility, (9), or belief sets, (10), are inter-definable: $wRv \equiv v \in Bel(w)$.

One of the first tasks of doxastic logicians has been to lay down axiomatic or semantic restrictions on $R$ (or, equivalently, on $Bel$) in order to refine the logical entailment relation to one that does justice to our intuitions about the logic of beliefs (Hintikka 1962). Other such accessibility restrictions can then be given for modalities other than belief. Epistemic logic, for example, models an agent’s knowledge and can be characterized by adding to doxastic logic an axiom to the effect that every proposition known is actually true. Some further generalizations of this general framework are multi-modal systems, for describing, say, the attitudes of both knowledge and belief in the same logical system, capturing their interrelatedness (e.g. knowledge implies belief), and multi-agent systems, for modeling interactions between multiple agents with different beliefs.

1.1.2.2 Possible worlds and representation

Let’s try to be a little more explicit on what it means for a world to be accessible or, as we will prefer to put it from now on, to be in the subject’s belief set. Haas-Spohn’s (1994:32) test offers some clarification:
Imagine the actual epistemic state of a person as fixed and then place him in a world $w$ which he may investigate in each and every detail. If he then finds no clues that $w$ is not the actual world—if, in other words, he can in no way distinguish $w$ from the real world as he knows it—then, and only then, will $w$ be an element of his belief set.

This characterization of doxastic alternatives thus relies on the observation that human beings' limited knowledge of the world can be captured by saying that we are not always able to tell two different possible worlds apart. In one extreme we have the omniscient whose belief set is a singleton containing only the actual world, so the omniscient creature thereby would believe all and only those propositions containing the real world, i.e. all true propositions. Other extremes are $\text{Bel}(w) = \emptyset$, the inconsistent believer who believes absolutely everything, and $\text{Bel}(w) = W$, the skeptic who believes only tautologies.

A belief set, like a proposition, is a set of possible worlds and as such can be seen as a representation of the world, and, consequently, as a model of the mental state of an individual representing the world as she perceives it. To clarify this, believing a proposition has been likened to dividing logical space (the set of all possible worlds) into those possible states of affairs the believer deems in accordance with how the world appears to her, and those she doesn’t believe to be possible, given her current cognitive state. For example, I wake up, the curtains are drawn, so I have no idea what the weather is like. In this case my belief set contains worlds in which it is raining (at this particular point in space-time) as well as worlds in which it isn’t raining but sunny, or snowing. By opening the curtains and looking outside I quickly conclude that it is in fact raining. My coming to believe the proposition that it’s raining amounts to repartitioning logical space so as to keep all rainy worlds in my belief set while discarding the sunny and the snowy worlds. This, basically, forms the intuitive basis of doxastic logic as a model for describing attitudes.

Simple and attractive though this picture may seem, it is not the only way to formalize a philosophy of attitudes. There are roughly speaking two trends in the philosophical literature on attitudes: possible worlds and propositionalism in the one corner, versus strong representationalism and mentalism in the other. The doxastic logic approach advocated so far falls squarely within the first. One of the main proponents of this view on attitudes, and cognition generally, is Stalnaker (e.g. 1984; 1999). Stalnaker claims that the explanation of human behavior and cognition must be sought at the abstract
level of information, as captured by possible worlds semantics. On such a view, meaning and mental states are not concrete, let alone structured entities, but rather abstract partitions of logical space, i.e. propositions, in the intensional logic jargon. This is what sets apart Stalnaker and his anti-mentalist company from so-called strong representationalists, among which we might group oldschool mentalists like Fodor (1975), but also semanticists like Asher (1986) and Kamp (1990), who claim that the objects of belief are highly structured entities, rather than abstract intensional objects.

Below, I briefly go into what I consider the chief argument against possible worlds approaches and in favor of (strong) representationalism. Nonetheless, I will eventually conclude that for the purposes of this work possible worlds will do just fine.

1.1.2.3 Logical omniscience and Frege’s puzzle

A possible worlds analysis of attitudes predicts that, if an agent believes a proposition $p$, she also believes every logical consequence of it, that is, any proposition $q$ that is less informative ($q$ contains all the worlds of $p$ and perhaps more). This defect of possible worlds semantics is often called the problem of logical omniscience and can be formulated as:

\[ (11) \quad \text{If } a \text{ believes } p \text{ and } p \subseteq q, \text{ then } a \text{ believes } q \]

In particular, it follows from (11) that nobody can believe Falsum (a common name for the empty set of worlds) and everybody automatically believes Verum (that is, the ‘always true’ proposition $W$). Actually, these things only become problematic when combined with the (usually implicit) assumption that, if an agent believes a proposition $p$, she is disposed to assent to any, or at least some, utterance expressing that proposition. A bi-conditional version of this principle is made explicit and named ‘the strengthened disquotational principle’ by Kripke (1979):

\[ \text{A normal English speaker who is not reticent will be disposed to } \]
\[ \text{sincerely reflectively assent to ‘p’ if and only if he believes that } \]
\[ p. \quad \text{[Kripke 1979:411]} \]

In 1.2.1 we return to Kripke’s disquotational principles in connection with \textit{de dicto} belief reports. For now, note that if we take the strengthened principle to mean that one believes a proposition iff one is disposed to assent

\footnote{Note that Stalnaker talks about fruitful ways of \textit{explaining human behavior}; he does not say that the actual attitudes occur independently of processes in the brain, described biologically or symbolically.}
to an utterance expressing it, we get from believing Verum to necessarily assenting to every expression of it, e.g. everybody who assents to something believes some proposition and therefore must believe its superset Verum. This proposition Verum is in turn expressed by a mathematical truth such as \textit{102,334,155 is the 40th Fibonacci number} to which perhaps not everybody would assent, given that even “sincere reflection” cannot always prevent miscalculations.

Another immediate and equally unwanted consequence of (11) in combination with the strengthened disquotational principle is the fact that co\textit{-intensional} terms are predicted to be always substitutable \textit{salva veritate}. This problem is often referred to as \textit{Frege’s puzzle} but I will argue that the common formulation of Frege’s puzzle—like logical omniscience—is actually a puzzle about attitude \textit{reports}. What I mean here is brought out by considering the pair of sentences in (12):

\begin{enumerate}
  \item The 40th Fibonacci number is greater than \(10^8\)
  \item 102,334,155 is greater than \(10^8\)
\end{enumerate}

I leave it to the reader to check that the 40th Fibonacci number is indeed 102,334,155, but assuming it is, it is necessarily so, since trivial truths of arithmetic like this are not contingent, i.e. if it’s true, it’s true in all possible worlds. In yet other words, the definite description in (12a) and the first number in (12b) are co\textit{-intensional} terms, meaning they refer to the same mathematical object in all possible worlds.\footnote{A Fibonacci number is the sum of the previous two Fibonacci numbers, given the starting values 0 and 1, i.e. \(F(n + 2) = F(n + 1) + F(n)\), \(F(0) = 0\), \(F(1) = 1\).}

In possible worlds semantics then, we must conclude that utterances of (12a) and (12b) express the same proposition, by an appeal to compositionality (semantic values of the parts, and their modes of combination are the same, so the sentences themselves are equivalent). Believing the proposition that the 40th Fibonacci number is greater than \(10^8\) is then the exact same thing as believing the proposition that 102,334,155 is greater than \(10^8\), a highly counter-intuitive prediction especially since, by the strengthened disquotational principle, we predict that everybody who assents to (12a) will \textit{immediately} assent to (12b) as well. Note that the derivation of this unwelcome result does not appeal to logical omniscience or its logical cause (10). The problematic immediacy is due to the fact that the disquotational principal as it stands does not take into account the process of determining what proposition is expressed by a given sentence in a context. In fact, this defect

\footnote{Cf. appendix 1.A for formal definitions of \textit{intension} and \textit{extension} in terms of reference and possible worlds.}
has been argued by many to be exactly the reason modal logic breaks down with respect to attitudes.

By placing more structure on the objects of belief, e.g. by considering sentential objects or structured propositions (Lewis 1972; Cresswell 1985; Soames 1985; Salmon 1986) we can block some unwanted omniscience and equivalence predictions. For example, simple intensional predicate logic cannot distinguish the propositions expressed by the sentences in (13): both are true in exactly the same possible worlds, viz. those where Ligia is a philosopher.

(13)  

a. Ligia is a philosopher  
b. 37 is prime and Ligia is not not a philosopher

Obviously, the sentences themselves are very different, so pure sententialism that says that belief is a relation between an individual and a sentence of her language has no trouble explaining this. The real challenge for such naïve sententialism is explaining how less articulate beings, like animals, can have attitudes and how people who speak different languages can nevertheless express the same beliefs.

Between these extremes of sententialism and propositionalism we have structured propositions. The idea behind it is that there’s a structural difference between the sentences in (13) and this may serve to tease apart attitudes towards their content. Structured propositions can successfully counter a lot of logical omniscience predictions, without suffering the sententialist defects: recognizing structure is something we may impute to most attitude holders, and language specific differences are neutralized because literal translation presumably preserves logical structure. The amount of structure put into propositions differs per theory, but we could imagine that the structured propositions expressed by the sentences in (13) are (14a) and (14b) respectively:

(14)  

a. \langle[Ligia], [philosopher]\rangle  
b. \langle\langle[37], [prime]\rangle, [and], \langle[\neg], \langle[\neg], \langle[Ligia], [philosopher]\rangle\rangle\rangle\rangle\rangle

With standard tools from type theory, it’s quite straightforward to retrieve also the classical propositions from these structured ones as well, so we could still make sense of the fact that, in some sense, these two sentences do mean the same thing.

Other strategies that have been pursued for endowing propositions with more structure involve tinkering with possible worlds, making them more fine-grained and/or allowing ‘impossible’ worlds (Montague 1970b; Cresswell 1972; Hintikka 1975; Rantala 1975). An alternative solution is called ‘awareness logic’ (Levesque 1984; Fagin & Halpern 1985), which divides a person’s
beliefs into explicit and implicit ones. The basic idea there is to allow a small (syntactically characterized) set of explicit beliefs while relegating all their infinitely many logical consequences to implicit belief. An overview of these semantic tweaks to classical modal logic can be found in Fagin et al. (1995).

But even within straightforward propositionalism, Stalnaker (1978) has devised a powerful way of exorcising logical omniscience. He maintains that belief objects are propositions and that propositions are just sets of possible worlds, but these propositions are not always just about objects having properties in the outside world but rather can be about the utterance or other linguistic structures that were used to convey them. This is formalized by means of the operation of diagonalization which turns the proposition that 37 is prime into the proposition that 37 is prime expresses a truth. We’ll return to discuss this operation in detail later, but note already that it could be useful for explaining the difference between Ligia saying “I am smart” and “That woman is smart” while pointing at a picture of herself. The same proposition is expressed, but there are many counterfactual situations in which uttering “I am smart” would express a truth while “Ligia is smart” would still be false.

What I hope to have shown in this subsection is that, although far from solved, the challenges posed by logical omniscience and Frege’s puzzle have been met with a wide variety of (partial) remedies. I will refrain from picking and implementing any of them in the following, in favor of a simpler, more elegant semantics. Though primarily a logico-philosophical problem, note that, due to the disquotational principle, omniscience problems have repercussions in the empirical linguistic domain concerned with the semantics of attitude reports. So, although in general, I try to separate these two fields of study as much as possible, here we see a case where a problem of belief logic infects the semantics of belief reporting. My view on the belief-report-interface is thus in line with the philosophy of mind position (as championed e.g. by Kamp (1990)) that holds that a study of the semantics of attitude reports presupposes an analysis of what attitudes themselves are. This contrasts with the opposite trend, common in analytic philosophy (cf. e.g. Dummett (1973)), according to which we can only learn anything about mental states and attitudes by first studying how natural language is used to talk about a person’s attitudes. My justification for using vanilla possible worlds semantics anyway is that the main issue I’m concerned with (the semantics of de re/de se reporting) is independent of logical omniscience in the sense that once an adequate omniscience free logic of belief is discovered, we should be able to just plug that into the semantic analyses described in the later chapters of this thesis.
1.1 Introduction: semantics and the attitudes

1.1.3 Modes of attitude: de dicto, de re, de se

The philosophical literature since medieval times distinguishes at least two distinct modes of attitudes, de re and de dicto. Briefly, a de re attitude is an attitude a person has towards some external object she is somehow acquainted with. A de dicto belief is just an attitude toward some general proposition. A little more recently, philosophical theories and observations about irreducible first person perspective (or essential indexicals) by Castañeda (1966), Perry (1977) and Lewis (1979a) among others, has led to the recognition of a third kind of attitudes, de se, paradigmatically comprising the beliefs a person has about herself based on her own unique first person point of view.

This subsection will discuss these three modes of belief in a little more detail, but on a pre-theoretical level, describing the phenomena and giving examples to highlight the differences between them. Theoretical accounts of the modes of attitude will be given in 1.2.

1.1.3.1 De dicto

Janell Ventress, who has a somewhat naive trust in academic justice, sincerely utters, “The best abstracts will be accepted”, thereby expressing that she believes the proposition that the best abstracts will be accepted. Let’s assume she hasn’t seen any of the abstracts. She did not mean to refer to some abstracts in particular, she was merely trying to convince her audience of the justness of the reviewing process, in other words, she meant that the acceptation of abstracts for this conference will be based on quality. On the basis of such a story we say that Janell believes de dicto that the best abstracts will be accepted.

De dicto or ‘notional’ (Quine 1956) belief is the paradigm mode of believing; it occurs when people believe a so-called general proposition. Typically we notice that someone has such a belief if they assent to, or themselves utter, a quantified statement expressing a generality. Typical general statements in this sense would contain quantifiers like every or a in some cases combined with ‘indefiniteness enhancers’ like whoever they may be which are typically used by semanticists to exclude de re readings. In the above anecdote, we can bring out this generality by following the Russellian treatments of definite descriptions, taking the definite article to be some kind of quantifier, but the important thing is that in this context it’s obvious that the statement does not predicate a property of some particular entity.

The term notional belief points to an alternative characterization: a belief ‘about concepts’ (notions) is de dicto, one that is about actual entities is de
Chapter 1. Attitudes

In the above scenario Janell’s utterance is not about particular entities, but can be analyzed as relating the concepts of acceptance and best abstract, stating that the second is subsumed under the first.

1.1.3.2 De re

As a member of the program committee, Janell has now been reviewing a number of abstracts. She thinks #037 is particularly insightful, and she’s convinced that the other reviewers will agree and that it will be accepted for presentation. In such a case we say that Janell has a de re belief: she believes de re of abstract #037 that it will be accepted.

The first defining characteristic of a de re belief is that it is about a particular object (res), not just a general belief about concepts. Furthermore it must be a particular object with which the believer has come into a reasonably direct contact, i.e. she must be acquainted with the res of her belief. Can we still analyze this as a propositional belief? Can we say that Janell believes the proposition that abstract #037 will be accepted? If so, note that this would impose some restrictions on the notion of proposition, viz. they must construed so that they can be said to be about things. But then, in what sense are mere sets of worlds about particular things? Can we model the object of Janell’s belief by the proposition that the thing called ‘abstract #037’ belongs to the set of things that will be accepted? What if in some worlds Janell’s abstract had been numbered #042?

The answer to these questions may be found by holding on to our starting point that in de re belief it is the object that counts, not its current look or the name it is endowed with in this particular world. If we wish to construe the object of Janell’s de re belief as a proposition, it must be a proposition about the abstract, not it’s name or the typeface or line spacing in which it is set at this or that world. This leads to the introduction of singular or Russellian propositions, i.e. propositions ‘containing’ objects themselves, in abstraction of their accidental properties. With the notion of a singular proposition (to be explicated in 1.2.2), presumably, the intentionality of de re belief is accounted for, but pondering the consequences the question soon rises, haven’t we gone a bit too far? Do human beings actually have beliefs about bare objects in abstraction of their modes of presentation?

A definitive negative answer to this question was provided by Quine (1956) who argues along the following lines: Imagine Janell’s hero is professor Ortcutt, she has read most of his articles and has even met him at conferences. Unbeknownst to Janell, Ortcutt submitted an abstract tentatively exploring his latest ideas. Janell is given this abstract for review, finds the author’s claims rather wild and unfounded, and consequently writes a
1.1 Introduction: semantics and the attitudes

devastating report. It appears that Janell has two very different de re beliefs, which happen to be about the same person, even though she herself doesn’t realize that. Of course, from the story it follows that Janell believes de re of Ortcutt that he’s a genius, but, intuitively independent of that, through reviewing his abstract she has formed another belief that is arguably also de re about Ortcutt, viz. that he is incompetent. The question of course is how to reconcile these two beliefs without ascribing her an inconsistent belief set, i.e. the logical analogon of total insanity, because that’s too harsh; she is merely confused about a factual identity of two personas, not insane. Many solutions have been proposed and most if not all of them are based on individuating de re beliefs more finely, somehow taking into account the way the res is presented to the believer on the occasion leading to the de re belief: in Janell’s first belief as Professor Ortcutt, the esteemed professor who wrote “Towards a Theory of Binding”, in the second as the guy who wrote this here lousy abstract. Note already that in doing so we appear to abandon the simple, binary, singular proposition approach.

Inspired by Quine’s analysis of Janell’s predicament we can extend the defining characteristics of the de re mode of believing: First, a belief can only be de re about an external object if the believer has somehow come in contact with that object. In the terminology I will be using: we require the believer to be acquainted with the res of her de re belief. Second, a de re belief always takes place under a description, viz. the believer’s relation of acquaintance with her res. Whether this leads to an analysis of belief as essentially a $\geq 3$-ary relation (believer, res, description/acquaintance relation, content of belief) remains to be seen. We will return to these issues of aboutness, singular propositions, modes of representation and acquaintance in a more theoretical discussion in 1.2.2.

1.1.3.3 De se

In order to introduce de se attitudes we must contrast them with mere de re belief, and for that purpose we complicate the story somewhat. As it turns out, the conference at Janell’s department is not exactly the best organized workshop ever, and not that fair either, since members of the program committee were allowed to send in abstracts of their own. In fact, Janell

---

7In Quine’s original, Ralph suspects a certain mysterious figure “glimpsed under questionable circumstances” of being a spy. At another occasion Ralph saw a gray-haired man at the beach “vaguely known to Ralph as rather a pillar of the community”. Both personas are in fact the same Ortcutt, so Ralph appears to have two conflicting beliefs about the same individual (Quine 1956:179). In the following I will often refer to the Janell paraphrase as Quine’s scenario.
herself had done just that. Furthermore, and this is crucial and typical of scenarios used to explain the de se mode, she forgot what her abstract was about—she sent it in really early, she’s been extremely prolific lately, she has very bad memory, she carelessly submitted a random old abstract without looking at it—and didn’t even recognize it as her own when she, by mistake or sloppiness of the organization, got it on her desk for review. The number used to identify this anonymous abstract of Janell’s was ... #037. The rest is history: she likes it a lot, gives it high marks, and trusts that it will be accepted for presentation. Thus we conclude that Janell believes de re of her own abstract that it will be accepted, and about herself that she will get to present.

A week after turning in the report, Janell is browsing through her old ~/Documents folder when she comes across an abstract which looks surprisingly similar to the one she has just reviewed. Realizing her mistake, she decides that it’s too late to bother the rest of the program committee now, and anyway, it was hardly her fault, was it? And her review was totally fair ... In short, she doesn’t change her belief that the abstract should be accepted, so she still believes de re of herself that she should be accepted. Something essential however has changed in this belief: we give this difference a name by saying that the belief is no longer merely de re, it is now de se.8

In 1.2.3 we will try to pin down these differences between de re and de se formally, but there are a few things we can point out here already. As discussed above de re belief requires a causal and cognitively effective connection between the believer and the res, the relation of x reviewing y for instance definitely counted as such an acquaintance relation between Janell and her abstract. Before realizing her mistake, Janell might have uttered something like “Abstract #037 is great” and would nod her head in agreement if she’d overhear someone say: “That abstract Janell’s reviewing will surely be accepted,” but of course she would never use the term my abstract9 to refer to her abstract until after her realization that the abstract was indeed hers. The de se modes requires a more particular kind of acquaintance, viz. of a first person type. In other words, in the final (de se) scene she is not only acquainted with her abstract as the abstract labeled #037, but as my abstract; same believer (Janell), same res (Janell’s abstract, aka #037), same belief content (that it will be accepted), but different descriptive presentation of the res.

Approaching this distinction from a slightly different angle, note that the my in the de se description of acquaintance cannot be substituted by a more

---

8This use of terminology is common, but later we need to refine it.
9Ignoring the reading of my abstract as the abstract I’m reviewing.
1.2 Modes of attitude

In the previous section we have looked at propositional attitudes, reports, and the modes of attitude (de dicto, de re and de se) on a level as pre-theoretical as possible. In this section we discuss some classic attempts at formally representing attitudes, capturing the distinctions between the three modes of attitude encountered in 1.1.3. For this purpose we’ll use first order modal predicate logic, with a possible worlds semantics, as defined properly in 1.A.

As promised before, I have tried to tease apart theories of attitudes and attitude reports: The current section explores what I classify as theories of attitudes, while chapter 2 deals with reports. Our present concern comprises philosophical theories about when a belief can be said to be de re about some some external object, or, say, the role played by perspective or point of view in beliefs about oneself, and how to formalize it all. The study of the semantics of attitude reports on the other hand aims at capturing the truth conditions of a certain class of natural language sentences, viz. those sentences that say that someone has some attitude to some information, or for the sake of the current work, sentences of the form NP believes that NP VP. However clear if presented in this fashion, the philosophical literature almost systematically fails to acknowledge this distinction. The connections
and mutual dependencies between the two being quite obvious indeed: the report “Colby believes the winner will be famous” is true under the \textit{de dicto} reading iff Colby believes \textit{de dicto} the proposition that the winner will be famous. It seems that first of all the semantics of belief reports depends on an analysis of beliefs, in the sense that we know the truth conditions of reports if we know how to represent what it is to believe a proposition. In 1.1.2.3 this was identified as a typical philosophy of mind position, diametrically opposed to the common analytical philosophy position. What I’m advocating is to minimize the amount of philosophical analysis about cognition and the intentionality of mind in order to focus entirely on giving linguistically satisfactory logical representations of belief reporting sentences in a given empirical domain. In order to define the important issues of \textit{de re}/\textit{de se} reporting we will have to lay some philosophical foundations in this chapter, and then, I contend, many linguistically interesting issues in report semantics can be investigated independently of pressing issues in belief logic, like logical omniscience.

\subsection*{1.2.1 \textit{De dicto} attitudes: the disquotation principle}

In the current section we take a closer look at \textit{de dicto} attitudes, in search of a definition of this type of belief. As a first approximation we noted in 1.1.3 that \textit{de dicto} is the basic mode of simply bearing the attitude of belief to a proposition, in contrast with believing something \textit{about} some external object (\textit{de re}), or about oneself from a first person point of view (\textit{de se}).\footnote{A well-known definition of the \textit{de re}/\textit{de dicto} distinction is based on a substitution principle: in \textit{de re} belief substitution of co-referential terms is OK, but in \textit{de dicto} it fails. However, this is a definition of \textit{de dicto} belief reports, at best. Moreover, as we will see in chapter 2, substitution of co-referential terms fails even in \textit{de re} reports.} \textit{De dicto} belief is the paradigm propositional attitude; it is the belief mode that is most straightforwardly described in simple modal logics of belief in possible worlds semantics. In fact, we will see it has been argued that, the other modes, \textit{de re} and \textit{de se}, are strictly speaking not even \textit{propositional} attitudes. \textit{De re} belief, for example, has been construed as relating a believer, an object, a property or open proposition, and a mode of presentation, whereas \textit{de se} belief is often explicited in terms of a relation between a person and property.

Only the \textit{de dicto} mode of believing then is adequately captured by definitions like (10), p. 9, which we now adapt to a definition of \textit{de dicto} belief:

\begin{equation}
(15) \quad a \text{ believes}_{\text{de dicto}} p \text{ at } w \iff \text{Bel}(w) \subseteq p
\end{equation}
Now, we would like a more naturalistic, theory neutral explanation of the phenomenon of de dicto belief. The key to such an explication, perhaps, lies in Kripke’s disquotational principle (cf. 1.1.2.3). Haas-Spohn (1994) uses the disquotational principle as a definition of de dicto belief ascriptions:

We will call a belief ascription de dicto if it can be reduced to an application of the disquotational principle or its extension: a statement of the form “a believes that ϕ” is a de dicto belief ascription if and only if it says that a would assent to ϕ (or an utterance of a translation of ϕ into a’s language). [(Haas-Spohn 1994:33)]

Observing the self-imposed constraint of strictly separating belief from belief reports I propose to take the principle as primarily a definition of de dicto belief, not de dicto reporting. Of course, de dicto reporting will be defined later in terms of de dicto belief, but this extra step counts and is relevant, especially in connection with de re and de se, where the linguistic surface forms of reports do not clearly match the corresponding belief modes. Moreover, I think my order of explanation is more compatible with Kripke’s original formulation:

If a normal English speaker, on reflection, sincerely assents to ‘p’, then he believes that p.  
[[Kripke 1979:411]]

First note that as is, this seems rather wild, for Quine (1953:117) already points out, such use of quotation marks does not make sense: who would assent to an utterance of the 16th letter of the roman alphabet? The best paraphrase I can come up with in our notation is:\footnote{Kripke (1979:248) himself adds that “‘p’ is to be replaced, inside and outside all quotation marks, by any appropriate standard English sentence”}

\begin{equation}
(16) \quad \text{If } a \text{ is disposed to assent to an utterance of } \sigma, \text{ then she believes the proposition expressed by } \sigma
\end{equation}

The type of belief that we can characterize thus, is precisely de dicto believing.

One thing worth noting here is that (16) only tells us what it means to believe a proposition de dicto if that proposition is expressible in a sentence of a’s language. This is no serious limitation for us however, because in the end, we are not studying the attitude of believing generally; all this is to be applied to the analysis of belief reports later on, and more often than not we will only report a person’s belief if that person has, on occasion, put her thoughts into words (or would do so, when prompted). On the other hand it may be argued that it is common practice to ascribe some forms of belief
to animals that clearly have no real language, and it is well-known that we often talk about the behavior of machines and computer programs as if they have beliefs, goals and intentions, if we want to describe their behavior. So perhaps we really should be able to cover those non-linguistic beliefs as well, especially when it comes to reporting. This view of belief is championed by e.g. Stalnaker (1984), who would consequently denounce any explications of de dicto or other belief in terms of the attitude holder’s linguistic abilities. In the following I will ignore this issue and restrict attention to believers that can put their beliefs into (English) words.

So, assuming you are willing to accept this restriction on the scope of our belief analysis, let’s return to defining de dicto in terms of a disquotational principle. Given the above discussion, we might want to include the expressibility clause, which would get us Kripke’s so-called strengthened biconditional version, which reads more like a real definition:

\[(17) \quad a \ \text{believes}_{de \ dicto} \ p \ \text{at} \ w \ \text{iff} \ w \ \text{there is a sentence} \ \sigma \ \text{of} \ a \text{’s language s.t.} \ \sigma \ \text{expresses} \ \text{the proposition} \ p \ \text{and} \ a \ \text{is disposed to assent to an utterance of} \ \sigma\]

In any case, the disquotational principle is really the only positive, non-circular characterization of de dicto believing. The de re mode, to which we turn next, has been given far more attention, leaving de dicto with a purely negative characterization as the mode of attitude that is not de re.

### 1.2.2 De re attitudes: singular vs. descriptive content

Whereas we analyzed de dicto belief in terms of a person standing in a certain relation to a proposition, the first thing to note about de re belief is that it seems, at first sight, to be more like a triadic relation, relating a believer, the object she’s acquainted with and which the belief is about (the res), and the actual content of the belief. Quine (1956:180) was the first to take this suggestion seriously and propose to analyze de dicto belief as dyadic and de re as triadic. Many others have wanted to avoid such a strict logical separation, especially since it leads to ambiguity in the analysis of some reports that can be read de dicto as well as de re (cf. 2.2 and (2) p. 72). The Ortcutt example of 1.1.3 shows we need to take into account yet another factor: the way the res is presented to the believer. In any case, the crucial role played by the res has prompted us to speak of someone believing something de re about something. This ‘aboutness’ or intentionality, in the sense of ‘being about a particular external object’, is the prime characteristic of de re belief.
1.2 Modes of attitude

1.2.2.1 Singular propositions

Singular propositions, we said, are propositions about actual individuals. Let’s take a closer look at their formalization in modal predicate logic. We assume a very simple ontology in which there are just objects, and these objects may have different attributes in different possible worlds. In such a framework we can formally distinguish objects and their accidental properties, among which the object’s name and shape. This is all we need to capture the notion of a singular or object-dependent proposition: a proposition about an object. Very literally, a singular proposition could be defined formally as a pair consisting of an object and a property, yielding in effect a structured proposition. We have seen those before (1.1.2.3), so recall that, for instance, the pair \( \langle P, a \rangle \), with \( a \) an object and \( P \) a property (set of objects), would represent the singular proposition which states of object \( a \) that it has property \( P \). This seems to be more or less what Russell had in mind, as witness for example his famous remark that Mont Blanc itself could well be a constituent of a singular thought, in the passage denounced by Frege (cf. correspondence with Russell (1904:169)); hence the term Russellian proposition for structured singular proposition. Using structured lfs for de re belief, as demonstrated in (18b), gets us essentially Quine’s aforementioned triadic de re analysis, (18a):

\[
\text{(18) } \begin{align*}
\text{Janell believes}_{\text{de re}} \text{ of abstract } \#037 \text{ that it will be accepted } \\
\text{a. } \triangleright \text{ BEL}(j, a, \text{accepted}) \\
\text{b. } \triangleright \text{ BEL}(j, \langle a, \text{accepted} \rangle)
\end{align*}
\]

We will not go any deeper into the theory of triadic belief or structured propositions here, since for our purposes it’s not necessary to change the basic setup of modal logic and possible worlds semantics. In the following I will try to show that we can simply stick with the by now familiar notion of a proposition as a set of worlds and a first-order language where belief is simply a modal operator (parametrized for believer, notation: \( \text{BEL}_j \varphi \) instead

\[12\text{Cf. (Neale 1990:49-50, note 1) for a discussion about the term singular proposition and the way it has been formalized by Russell. My above attempt at a theory neutral characterization of singular in terms of aboutness deviates slightly from Neale’s definition as a proposition expressed by a sentence containing a referential term. I merely attempted to abstract away from Neale’s (and others’) reliance on the level of linguistic/logical semantics. The reason being that his definition runs a significant risk of circularity if at some point we want to characterize a certain class of linguistic constructions by saying that they express singular propositions.}\]


\[14\text{The link between structured lfs and the structured propositions encountered in (14), p. 13, is, of course: } \langle \alpha, \beta \rangle = \langle \llbracket \alpha \rrbracket, \llbracket \beta \rrbracket \rangle\]
Chapter 1. Attitudes

of $\text{BEL}(j, \varphi)$. At first sight, it seems this gives nothing but a new notation of the dyadic, de dicto, belief relation. And indeed it does that, as shown below:

(19) Janell believes, abstract #037 will be accepted

$\vdash \text{BEL}_j \text{accepted}(a)$

$\llbracket (19) \rrbracket_w = 1$

iff in all $w' \in \text{Bel}(\llbracket j \rrbracket, w)$: $\llbracket \text{accepted}(a) \rrbracket_{w'} = 1$

iff in all belief alternatives $w'$ (of Janell in $w$), the abstract #037 of $w'$ gets accepted in $w'$

Crucially this is not a belief about abstract #037, the actual object, but about the abstract numbered #037 whatever it may turn out to be. The difference becomes apparent, for instance, if we assume that there is actually no abstract #037, but, before the call for papers, Janell has already formed the opinion that every abstract will be accepted, because there aren’t going to be enough submissions anyway and the department will agree that refereeing is just too much work. The disquotational principle would yield that Janell holds a de dicto belief towards the proposition that abstract #037 will be accepted, and sure enough, in a model based on such a case (19) would be true. On the other hand there’s no acquaintance with actual abstracts, so no de re belief. The other way around, imagine that Janell has mixed up #037 and #042, i.e., in her belief worlds the term abstract #037 refers to the actual #042, and vice versa. In fact, she read #037 and found it excellent, but, in her confusion, has told the committee, sincerely, that “#042 should be accepted, #037 is awful.” Obviously this does not constitute a positive de dicto opinion about (what she considers to be) abstract #037, and, accordingly, (19) comes out false, but there is an actual acquaintance relation between her and #037 on the basis of which she forms her opinion, so that we’re entitled to say that she believes de re of abstract #037 that it will be accepted (even though she mistakenly thinks it’s labeled #042).

So much for de dicto in modal logic. Let’s see how this framework is able to formalize the idea of a singular proposition as well. This can be done in two ways: wide scope and directly referential constants, to be discussed in succession. The scopal analysis of de re belief depends on something called ‘quantifying into’ a modal operator to represent de re belief:

(20) Janell believes de re of abstract #037 that it will be accepted [=\((18)\)]

$\vdash \exists x [x = a \land \text{BEL}_j \text{accepted}(x)]$

$\llbracket (20) \rrbracket_w = 1$
iff in all \( w' \in Bel([j]_w, w) \): \( \left[ \text{accepted}(x) \right]_{w'}^{x\rightarrow [s]} = 1 \)

iff in all belief alternatives \( w' \) (of Janell in \( w \)), the abstract #037 of \( w \) gets accepted in \( w' \)

The truth conditions computed above have given us something comparable to (19), but now the res-denoting term, abstract #037, or \( a \), is evaluated in the actual world \( w \), not one of the belief worlds \( w' \). This is enforced representationally by the fact that the res is represented by a variable whose value is fixed by the \( x = a \) clause outside the modal context. Technically, the computation thus relies on assignment functions \( (x \mapsto [a]_w) \), which are defined properly in 1.A. The crucial feature of the whole thing is that the predicate may have varying extensions across Janell’s belief world, but in the course of the semantic derivation the res represented by the bound variable \( x \) stays mapped to the actual abstract #037 irrespective of Janell’s beliefs about for instance its quality, author, number, or even existence.

The direct reference approach takes its cue from the above to present a shortcut: the res of a de re belief may simply be represented by an individual constant in situ so the logical form for Janell’s de re belief is simply (2), but to get the truth conditions of (20) we add the stipulation that constants have rigid intensions, i.e. for all \( w, w' \in W \): \( [a]_w = [a]_{w'} \). Assuming beforehand that abstract #037, \( a \), refers to the actual #037 in every possible world there is of course no longer any truth conditional difference between (19) and (20), so this works fine. At this point however it may seem a bit far-fetched to stipulate rigidity for all constants. Therefore, we’ll stick with the more conservative ‘quantifying in’ solution so that the de re nature of a report is immediately apparent from the lf. We return to rigidity later when we can give actual linguistic/semantic evidence for direct referentiality of many terms logically represented as constants (indexicals, proper names, natural kind terms, (some) definite descriptions). The term direct reference, which we will formalize as rigidity, points to the fact that, as was shown by Kripke (1972) and Kaplan (1989), these terms simply refer, not because their content singles out a unique individual in the world, but more like tags or labels attached to an individual.\(^{15}\)

1.2.2.2 Substitution

All in all, it seems that having singular propositions (formalized in terms of structured propositions, wide scope, or rigid designators) as objects of de re attitudes is the way to reconcile the intentionality of de re belief with

\(^{15}\)In addition we’ll be able to define Kaplan’s operator \texttt{dthat}, which rigidifies a given descriptive term so we don’t have to stipulate rigidity for all constants.
the traditional notion of attitudes as dyadic relations between persons and propositions. On such a construal we naturally obtain compatibility with the traditional definition of \textit{de re} modality as that for which we have \textit{substitutio salva veritate} of co-referential terms.\textsuperscript{16} With that definition, a belief report about abstract \#037 would only count as \textit{de re} if the report remains true if we replace the \textit{res} denoting expression \textit{abstract \#037} with a co-referential expression, like \textit{Janell’s abstract} or \textit{the abstract Janell is reviewing}. A dubious prediction for sure, but one concerning linguistic intuitions about reports, not the theory of attitudes.

However, taking predicate logic as a formal language to describe truth conditions we can mimic this result within the theory of attitudes proper, viz. by applying \textit{substitutio} at the level of logical form. Thus, a logical formula with a modal operator is \textit{de re} with respect to a certain term iff replacing that term with a co-referring one does not alter the truth value. Our definition of \textit{de re} belief as involving wide-scope (or rigid) \textit{res} representations verifies this result, since:

\begin{equation}
\begin{align*}
(21) \quad \text{For any terms } \alpha \text{ and } \beta \text{ with } [\alpha]_w &= [\beta]_w: \\
&\left[\exists x [x = \alpha \land \text{BEL}_j \text{accepted}(x)]\right]_w = \left[\exists x [x = \beta \land \text{BEL}_j \text{accepted}(x)]\right]_w
\end{align*}
\end{equation}

Even without dragging truth value judgments of belief reports into the discussion at this point, it is obvious that we are going to run into problems with this substitution result. To see this, consider again the example where Janell has two \textit{de re} beliefs about the same Ortcutt, under two different guises, viz. \textit{the famous professor Ortcutt} and \textit{the guy who submitted abstract \#042}. In predicate logic, we could try to model this situation by positing two individual constants, say \(f\) (amous professor) and \(g\) (uy who submitted abstract \#042). These constants represent individual concepts, i.e. they are not rigid per definition but may have rich descriptive contents: \([f]\) is something like the concept mapping each world to the famous professor admired by Janell in that world. This is to say that \(f\) and \(g\) are semantically equivalent to definite descriptions, and in fact we might as well replace all occurrences of these individual constants below with the corresponding \(\tau\)-terms \(\iota z [\text{famous prof}(z)]\) and \(\iota w [\text{submit abstr} \_42(w)]\).

Getting back to the Ortcutt case, the following facts are immediately obvious and uncontroversially true in the scenario:

\textsuperscript{16}Although that definition is most straightforwardly read as a definition of \textit{de re} belief reports, I mention it here because, as I’ve said before, the distinction between beliefs and reports has not always been drawn so sharply and this is a case in point, classic authors claiming substitutability as definition—or problematic result—for \textit{de re} belief.
1.2 Modes of attitude

(22) a. Janell believes_{de dicto} that the famous professor Ortcutt is a genius
\[\triangleright \text{BEL}_j\text{genius}(f)\]

b. Janell believes_{de dicto} that the guy who wrote abstract #042 is a nincompoop
\[\triangleright \text{BEL}_j\text{nincompoop}(g)\]

c. The famous professor Ortcutt = the guy who wrote abstract #042 = Ortcutt
\[\triangleright f = g = o\]

Moreover, as we—along with Quine (1956), Kaplan (1969) and others—have assumed before, the story justifies a strengthening of (22a) and (22b) to the corresponding \textit{de re} beliefs (23a) and (23b), respectively:

(23) a. Janell believes_{de re} of the famous professor Ortcutt that he is a genius
\[\triangleright \exists x [x = f \land \text{BEL}_j\text{genius}(x)]\]

b. Janell believes_{de re} of the guy who wrote abstract #042 that he is a nincompoop
\[\triangleright \exists y [y = g \land \text{BEL}_j\text{nincompoop}(y)]\]

These two together with (22c) entail (24a-b), which in turn make it impossible to block the inference to the inconsistent (24c), that is, in modal logic as we know it: (23a),(23b),(22c)\models(24a)\land(24b)\models(24c)

(24) a. Janell believes_{de re} of Ortcutt that he is a genius
\[\triangleright \exists x [x = o \land \text{BEL}_j\text{genius}(x)]\]

b. Janell believes_{de re} of Ortcutt that he is a nincompoop
\[\triangleright \exists y [y = o \land \text{BEL}_j\text{nincompoop}(y)]\]

c. Janell believes_{de re} of Ortcutt that he is a genius and a nincompoop
\[\triangleright \exists x [x = o \land \text{BEL}_j[\text{genius}(x) \land \text{nincompoop}(x)]]\]

In other words, the wide scope analysis cannot account for the fact that Janell’s beliefs are both \textit{de re} and, in a sense, about two subjectively distinct, but \textit{actually} and \textit{contingently} co-referential concepts, without thereby discarding Janell’s belief set as internally inconsistent. In the rest of this section we’ll try to come up with a way to block the problematic inference to an inconsistent belief state without denying that Janell’s beliefs are \textit{de re}.

For completeness, note that a rigid designator-based theory is even less suited to model the situation, since the assumed \textit{de re} nature of the beliefs would require the representations of the res, f and g, to be rigid and therefore
co-referential not only in the actual, but in all worlds, thus entirely missing the point of the two descriptions that seem to play a role in the story. The structured propositions theory in the simple form discussed above suffers from the same defect: by definition, it neglects modes of presentation of the res, thus eventually predicting inconsistent beliefs for Janell. It seems then that singular propositions, in all three forms discussed, are insufficient as objects of de re belief, we need modes of representation, and moreover, we need them to play a role inside the belief (as opposed to their place in the wide scope analysis).

1.2.2.3 Belief under a description and the shortest spy problem

We start by extending our predicate logical analysis to take modes of presentation into account, and then test the apparatus by applying it to our Ortcutt example. What we need is not simply \( x \) believes de re of y that so-and-so, but \( x \) believes de re of y under description \( z \) that so-and-so. The descriptions are intended to differentiate subjectively distinct de re beliefs, since a de re belief’s aboutness (formalized above in terms of singular propositions) only distinguishes beliefs with externally distinct objects. In short we add a fourth argument to the belief relation:

\[
\text{Bel} (j, a, \text{accepted}, d)
\]

In predicate logic we can use \( \iota \)-terms or constants denoting individual concepts to represent these descriptions. Let’s do the latter, and let’s distinguish these intensional constants/variables from simple referential ones at If by making them just a bit more curly, typographically. Holding on to the wide scope analysis of intentionality, we posit the following official notation:

\[
\text{Bel}(j, a, \text{accepted}, d) \iff \exists x [x = b \land \text{Bel}_d^a(P(x))]
\]

Eventually we will formally reduce this fourth argument slot of the belief relation to more primitive machinery, but, leaving it as a primitive for now, let’s see how it would help us in the Ortcutt case.

Above we already speculated on what would be the appropriate description: the first belief involves Ortcutt under the description the famous professor Ortcutt (\( f \)), the second takes place under the guy who wrote abstract #042 (\( g \)). The beliefs we had at the near final stage in (24) now come out as follows:
1.2 Modes of attitude

(27) \[ \exists x \in \mathcal{X} \mid x = \mathbf{0} \land \text{BEL}_0^\downarrow \text{genius}(x) \land \text{BEL}_0^\downarrow \text{loser}(x) \]

The intensional difference between these descriptions can perhaps serve to block the inference to an inconsistent \textit{de re} belief. Note that this seems plausible since intuitively there really is no single description of Ortcutt under which we can say that Janell believes he’s both a genius and a nincompoop. To be sure, we need to say a bit more about what \textit{belief under a description} is, and how it relates to \textit{de re} belief as such.

The most straightforward interpretation of descriptions in belief that comes to mind is the following: First, one way to define \textit{belief under a description} is to literally take descriptions into the belief, turning the \textit{de re} complements into truly \textit{de dicto} or general beliefs by replacing res with their intensional descriptions. A slight complication: for ‘type reasons’ we need a \( \lor \) operator to collapse the intensional object to whatever reference it currently has (\( \llbracket \lor d \rrbracket_w = \llbracket d \rrbracket_w(w) \)). The result:

(28) \[ \text{BEL}_a^\downarrow P(x) \sim \text{BEL}_a P(\lor d) \]

In words, Janell believes \textit{de re} of Ortcutt that he is a genius under the description \textit{the famous professor} iff she believes \textit{de dicto} that the famous scientist known as ‘Prof. Ortcutt’ is a genius (\( \text{BEL}_1^\downarrow \text{genius}(\lor f) \)). This seems to work fine for the current scenario where the so-derived \textit{de dicto} belief is indeed appropriate and true, as witness (22a). Secondly, we still have to define \textit{de re} belief. The most straightforward solution is existential quantification, i.e., \( a \) believes\textit{de re} of \( b \) that it’s \( P \) iff there is some description \( d \) of \( b \) such that \( a \) believes\textit{de re} of \( b \) under \( d \) that it’s \( P \). Obviously, an individual concept ‘describes an individual \( b \)’ in \( w \) iff \( d(w) = b \). Combining these insights with (28) gives:

(29) \[ a \text{ believes}_{\text{de re}} of b \text{ that it is } P \]
\[ \text{iff } \text{there is a concept } d \text{ describing } b \text{ such that } a \text{ believes}_{\text{de dicto}} of d \text{ that } d \text{ is } P \]
\[ \lor \exists d[\lor d = b \land \text{BEL}_a P(\lor d)] \]

We’ll refer to (29) as the \textit{de re} analysis by \textit{unrestricted quantification over concepts}. At first sight, it seems to work: Janell’s \textit{de re} beliefs about Ortcutt come down to two perfectly consistent \textit{de dicto} beliefs, viz., under \( f \), that the famous professor is a genius; and, with \( g \), that the author of abstract \#042 is a nitwit. In addition, we have successfully blocked the counterintuitive inference that Janell believes of Ortcutt that he is both, for no \( d \) satisfies contradictory predicates.
It turns out that this theory is a bit too liberal in assigning *de re* beliefs. This is shown by Kaplan’s (1969) so-called ‘shortest spy problem’ which received its name from a scenario where someone sincerely assents to the tautology: “The shortest spy is a spy.” As it happens, the shortest spy is actually Ligia, but nobody knows this, moreover, the speaker above has never even met or heard about Ligia. On the current theory, though, this person is predicted to have a *de re* belief about Ligia to the effect that she is a spy, taking *the shortest spy* as description of Ligia. Now, this is absurd, *de re* beliefs were supposed to be beliefs about specific individuals with which one is vividly acquainted. We conclude that the analysis based on unrestricted quantification over concepts, is defective: we need to take the acquaintance/intentionality aspect into account.

### 1.2.2.4 Acquaintance relations

A *de re* belief, we said, is a belief about a particular object in the outside world. Repeating the argument in 1.1.3, it follows that, in the case of a belief *de re* about an object, the believer must be *en rapport* with that object.\(^{17}\) This is where the notion of acquaintance comes in. Following Lewis’s (1979a) terminology, we say that *de re* believing requires the believer to stand in a proper relation of acquaintance to the *res*. This terminology is reminiscent of (and derived from) Russell’s *knowledge by acquaintance*, but our notion of acquaintance is not nearly as strict as the one he envisaged. In fact, according to Russell (at least in 1912:57), the only thing one has knowledge by acquaintance of is oneself, which is way too strong a restriction for a characterization of *de re* attitudes.

What then do we count as valid relations of acquaintance? Apart from the fact that they are two-place relations between individuals and other external objects there are, as far as I know, no hard and fast rules. Moreover, it really is a matter of considerable philosophical debate, since an answer would be intimately connected with metaphysical and epistemological issues, that is, what is the external world made of and how can we know anything about it? I certainly won’t draw the line, so let’s assume that we’re dealing with a vague notion here. This is not to say that there isn’t an identifiable subclass of prototypical acquaintance relations, and another of obvious no-gos. Typically, such direct perceptual relations as seeing, hearing, touching, etc. are in, while having heard that someone once claimed that *x* existed, or knowing only its name, are out. In between are indirect perception cases like reading

---

\(^{17}\)Cf. Evans’ Russell’s Principle, which states that “in order to have a thought about a particular object, you must know which object it is about which you are thinking” (Evans 1982:74).
1.2 Modes of attitude

an email from someone addressed to you, seeing someone on TV, in a mirror, while on drugs, or all at once. In most cases I would still count these as proper relations of acquaintance, but note that especially in these cases error is possible, in the sense that what we actually bear an acquaintance relation to does not necessarily coincide with the object we think we bear the relation to: we are not always aware of the actual objects of our acquaintance. This will turn out to be a crucial feature of acquaintance relations.

One interesting consequence of the vagueness of what counts as an acquaintance relation is that we seem to end up with a corresponding vagueness in the \textit{de re}/\textit{de dicto} distinction. The more direct and vivid the relation, the more justified we are in calling a belief based on it \textit{de re} about the so-acquainted \textit{res}. On the other hand, a distant, indirect relation to a \textit{res} will merely give rise to a \textit{de dicto} attitude. Since I deem a strict division of \textit{de re} and \textit{de dicto} attitudes highly implausible, I consider this vagueness, inherited from the notion of acquaintance, not a bug, but a feature of the analysis. One could even go one step further and give up the whole \textit{de dicto}/\textit{de re} distinction by allowing any relation to count as acquaintance. For example, we would then count in an acquaintance with someone as \textit{whoever will win the race}, which traditionally would only give you a \textit{de dicto} belief, since knowing that, by the rules of the game, there will be a unique winner does not qualify as the \textit{knowing who} needed for \textit{de re}-aboutness. This could perhaps be put to use in a unified account of \textit{de dicto} as \textit{de re}, a suggestion briefly taken up in chapter 3, p. 231. On further reflection, it becomes apparent that something extra is need or we’d end up with the previously dismissed unrestricted quantification over concepts (p. 30), in which there is a belief about something as soon as there exists some description so that the belief holds of the description. Something along the lines of a restriction to contextually salient or vivid concepts must be added to avoid the shortest spy problem. For now, we’re just interested in \textit{de re}, so we’ll add restrictions like vividness and (perceptual) acquaintance.

To sum up, a necessary prerequisite for \textit{de re} believing is a relation of acquaintance between believer and \textit{res}. Furthermore, relations of acquaintance can play the role of a mode of presentation of the \textit{res} for the believer, so it seems we have found a candidate descriptive content to be incorporated into the belief. This proposal comes down to analyzing \textit{de re} belief as a belief about an external object under the description given by the relation of acquaintance that constitutes the link between the believer and that external \textit{res}. We still have to make precise what \textit{belief under a description} means, but the idea will be clear from an example, so let’s briefly return to Janell and professor Ortcutt’s abstract #042. The thing to note is that there are two acquaintance relations at work, firstly, Janell is acquainted with Ortcutt as
the guy she saw on book covers and conferences, and it is this acquaintance relation that gives rise to the belief about Ortcutt being a genius. The other salient relation of acquaintance between Janell and Ortcutt here consists in her knowing the guy merely as the author of this lousy abstract. In short, it’s exactly these two ways of knowing the actual Ortcutt that give rise to the two distinct beliefs, that is, two beliefs about the same object, under two different guises. Taking for each belief the particular acquaintance relations that gave rise to, or is salient in, that belief as the description for it, should give the right result.

1.2.2.5 Kaplan (1969): reducing \textit{de re} to \textit{de dicto}

In a reaction to Quine’s (1956) postulation of an essentially triadic analysis of \textit{de re}, comparable to the analysis in terms of structured propositions, Kaplan (1969) tries to reduce \textit{de re} to a kind of \textit{de dicto}. Where Quine needed a strict ambiguity between two belief operators, a so-called notional one (\textit{de dicto}) and a relational (\textit{de re}), Kaplan argues we need only one, the good old dyadic propositional attitude operator, relating an individual to a (non-singular) proposition. In this sense Kaplan’s analysis is correctly characterized as a reduction of \textit{de re} to \textit{de dicto}. It is especially suited for our purposes, since in addition to being a conceptual reduction (as was unrestricted concept quantification), it is one that incorporates neatly the insights having to do with the importance of acquaintance and their relation to \textit{res} descriptions, as discussed in the previous subsections. So, let us now look at the details of Kaplan’s analysis by applying it to our Ortcutt example. We have already outlined our strategy: take mode of presentation into account, but not just any mode of presentation, just the ones that derive from a genuine relation of acquaintance.

Kaplan follows the unrestricted concept quantification account in reducing \textit{de re} belief under a description to \textit{de dicto} belief, i.e. \( \text{bel}_a \mathcal{P}(x) := \text{bel}_a \mathcal{P}(\gamma d) \) \[= (28) \]. The thrust of his theory of belief under a description however lies in the way the quantification over descriptions is restricted: a \textit{de re} belief requires a way of being acquainted with the \textit{res}, and this acquaintance relation has to provide the mode of presentation for the belief. In Kaplan’s (1969) terminology a \textit{de re} belief implies a vivid name of the object for the believer. To represent this, he posits a 3-place relation \( R(\alpha,b,a) \), \( \alpha \)

\footnote{Actually, in the cited paper, Kaplan favors a sententialist rather than propositional view of attitudes, but the difference is an issue orthogonal to the one at hand (recall the discussion of logical omniscience), and since we have already decided to go with the well-studied standard view of attitudes as propositional, and of propositions as sets of possible worlds, I will simply translate Kaplan’s proposal into that framework.}
1.2 Modes of attitude

represents \( b \) to \( a \). The \( a \) and \( b \) denote believer and \( res \), respectively, and his \( \alpha \) ranges over names, but may in our more semantic framework (cf. footnote 18) be taken to range over individual concepts, \( d \). According to Kaplan, \( \mathcal{R} \) really abbreviates 3 conditions:

\[
\mathcal{R}(d,b,a) \text{ iff }
\]
\[
(i) \ d \text{ denotes } b \\
(ii) \ d \text{ is a name of } b \text{ for } a \\
(iii) \ d \text{ is (sufficiently) vivid}
\]

Clause (i) says that \( d \) in the actual world describes the object \( b \). Clauses (ii) and (iii) require some extra attention. As for (ii), a name is of \( b \) for \( a \) iff \( b \) is causally linked to the name, by dubbing, and consequently to \( a \), by usage of that name. In other words \( b \) was present as target of the so-called initial baptism or dubbing event, after which a causal chain of events got started, which ultimately ends up connecting \( b \) to \( a \). Names in the Kaplanian sense however are not restricted to the linguistic class of proper names, descriptions like our the famous professor and the guy who wrote abstract \( \#042 \) can serve as names of Orctutt for Janell. In these cases there is no institutionalized baptism, but there Janell presumably uses that description as a mental or verbal tag for a specific individual she knows. Note further that on our more semantic interpretation, these ‘names’ and ‘descriptions’ we keep talking about are actually formalized as the intensions associated with such linguistic items. Clause (iii) says that the description should be vivid, which means that it should be associated, in the subject’s mind, with a halo of detailed information, descriptions, mental pictures, of the so named individual. As a rule of thumb, Kaplan offers the criterion ‘the more detailed, the more vivid’.

The full definition of \( de re \) belief, extending (29), then comes out as follows:\(^{19}\)

\[
(31) \quad a \text{ believes}_{de \text{ re}} \text{ of } b \text{ that it is } P \\
\text{ iff there is an individual concept } d \text{ s.t. } \mathcal{R}(d,b,a) \text{ and } a \text{ believes}_{de \text{ re}} \text{ of } b \text{ that it is } P \text{ under description } d
\]
\[
\text{ iff there is a } d \text{ s.t. } \mathcal{R}(d,b,a) \text{ and } a \text{ believes}_{de \ text{ dicto}} \text{ that (the) } d \text{ is } P
\]
\[
\text{ iff there is a } d \text{ s.t. } \mathcal{R}(d,b,a) \text{ and } a \text{ believes}_{de \ dicto} \text{ that (the) } d \text{ is } P
\]

\(^{19}\) \( \mathcal{R} \) denotes Kaplan’s \( \mathcal{R} \), which in turn should be spelled out along the lines of (30). Whether this can be done within a first-order system is irrelevant since the quantification over individual concepts (\( \alpha \)) already gets us beyond beyond first-order.
In words: a *de re* belief about *b* is defined as the *de dicto* belief in the general proposition that *d* has this or that property, for some appropriate descriptive name *d* that the subject uses to identify *b*.

For example, Janell believes *de re* of Ortcutt that he’s a genius because (i) there is a name *the famous professor Ortcutt* which names Ortcutt for Janell, which actually denotes Ortcutt, and which is vivid for Janell, and (ii) Janell believes *de dicto* that the famous professor Ortcutt is a genius. On the other hand she also believes *de re* of Ortcutt that he is a nincompoop, which is shown by taking the name to be *the guy who wrote abstract #042*. Lastly, note that there is no vivid name *α* of Ortcutt or anybody else s.t. she believes *de dicto* that *α* is a genius and a nincompoop at the same time. The only plausible candidate *α*, ≈ *the person who is either the famous professor or the author of the abstract* (cf. footnote 18), is not at all vivid, nor, arguably, a name of Ortcutt for Janell. This ‘frankly inegalitarian attitude toward various ways of specifying [an individual]’ (Kaplan 1969:222 & note 17), is the key feature in Kaplan’s solution to the Ortcutt problem with which we conclude the discussion of *de re* belief and its representation in modal predicate logic.

Hoping to have convinced you of the soundness of Kaplan’s *de re* reduction with acquaintance, on which the rest of the thesis will build, let me also raise a point of concern that, perhaps, need not bother us here already, but that will inevitably rear its head once we want to expand our analysis to attitudes beyond belief. Consider what happens if we just copy the entire analysis to an analysis of the propositional attitude of *hoping*:

\[
\begin{align*}
(32) \quad \text{a hopes}_{de \, re} \text{ of } b \text{ that it is } P \\
\text{iff} \quad \text{there is an individual concept } d \text{ s.t. } R(d, b, a) \text{ and a hopes}_{de \, dicto} \\
\text{that } P(d) \\
\quad \Rightarrow \quad \exists d[R(d, b, a) \land \text{HOPE}_a P(\forall d)]
\end{align*}
\]

The problem arises when the person you believe falls under description *d* differs from the person you hope falls under *d*. Say, Janell knows that Ortcutt, the famous professor, is head of Department X, but that she’s expecting him to retire soon. Her favorite candidate for Ortcutt’s succession is her good friend Ligia Faust. As before, she’s still a big fan of Ortcutt’s work and obviously hopes that Ortcutt’s abstract gets accepted. Because Ligia hasn’t submitted anything (or submitted something terrible), Janell does not hope that Ligia is accepted. In such a scenario, we are surely entitled to say that Janell hopes_{de \, re} of Ortcutt that he is accepted, under the description *my esteemed colleague and head of Department X*. Under Kaplan’s criteria, such a description would most likely count as a vivid name of Ortcutt for
Janell. However, having a *de dicto* hope that her esteemed colleague etc. gets accepted, means that in all worlds compatible with Janell’s hopes, the colleague that is head of Dept. X in that world gets accepted. And unfortunately that is false, since, according to the above story, in all of Janell’s hope-alternatives Ligia is head of Dept. X and, moreover, we assumed that Janell does not hope that Ligia gets accepted.

Apparently, the generalization to other attitudes than belief requires more than just a copy of (31) replacing each ‘believes’ with the attitude under consideration. What is needed is a way to account for the fact that attitudes like hope asymmetrically depend on belief in the sense that the *de dicto* hope in the definiens of (32) does not concern the individual hoped to be *d*, but rather the one believed to be *d*. To put it another way, belief is the basic attitude on which all the others depend, for example, hoping that Janell is home implies a belief set with a person named Janell, and, on top of that, the hope that that person in the belief worlds is home. Trying to make this idea formally precise, we quickly run into serious trouble, which is the main reason we’ll be almost exclusively concerned with the attitude of belief for the remainder of this work (but occasionally we will come back to the other attitudes). The aim is a theory of belief that is readily extendable to the other attitude, pending an analysis of this asymmetric belief dependence. I refer the interested reader to the works of Kamp and Asher that explore these interdependencies between attitudes in some detail (Asher 1987; Kamp 1990).

1.2.3 *De se* attitudes: indexical belief as self-ascription of properties

Since Casta˜neda (1966), and later Kaplan (1989), Lewis (1979a) and Perry (1977; 1979) it is generally granted that there is a third mode of believing; besides *de dicto* and *de re* there is also *de se*. The motivating examples for this mode are cases like Kaplan’s pants-on-fire, Perry’s lost amnesiac Lingens, Lewis’ propositionally omniscient gods, or our Janell who does not realize she’s reviewing her own abstract (p.17). The main argument is this: the defeatist belief expressed by *my abstract is no good* is different from any *de dicto* belief in a proposition that the such and such abstract is bad (because she may fail to realize that the such and such abstract is hers), or from merely a *de re* belief about Janell’s abstract to the effect that it’s no good (because she has that *de re* belief already on the basis of holding the abstract and looking at it, saying that it’s unacceptable, even when she hasn’t yet recognized it as her own).
In this section we take a closer look at these \textit{de se} examples, terminology and arguments, and we shall review the discussion about whether we really need 3 primitive modes, or whether we cannot reduce one to another. We’ll end up extending the basic possible worlds framework somewhat by saying that belief in general relates individuals to properties, not propositions, or equivalently, that belief alternatives are contexts rather than worlds (Lewis 1979a). Then we use Kaplan’s (1969) acquaintance relations (1.2.2) to reduce \textit{de re} to (generalized) \textit{de dicto}, and finally we’ll argue that \textit{de se} is just a subclass of \textit{de re} (Lewis 1979a; Boër & Lycan 1980; von Stechow 1982). The debate between \textit{de se} reductionists and separatists becomes more heated when we enter into the semantics of belief reports, but that will have to wait until chapter 2.

1.2.3.1 The essential indexical

Lewis (1979a) invented the term \textit{de se} in analogy with \textit{de dicto} and \textit{de re} beliefs, and we will use his terminology for what is also often called indexical belief. The latter term being perhaps more apt since \textit{de se} may suggest too strong a connection with strictly first person, or ‘self’, belief, i.e. beliefs one would typically express with the help of the first person pronoun \textit{I},

\footnote{We focus on English throughout, but especially in these philosophical sections.}

although at first sight there seems to be no good reason to set apart \textit{I} from other indexicals. So how do we characterize the \textit{de se} mode? Answer: a belief is \textit{de se} iff the believer could not have expressed it without indexicals,

\footnote{This is the second place where our discussion crucially depends on beliefs being expressible in a language. As said before, when discussing the disquotational principle in 1.2.1, it may be a significant omission, but I will restrict attention to this class of beliefs.}

although at first sight there seems to be no good reason to set apart \textit{I} from other indexicals. So how do we characterize the \textit{de se} mode? Answer: a belief is \textit{de se} iff the believer could not have expressed it without indexicals,

\footnote{And meaning here should be taken as a primitive, pre-theoretical notion, not necessarily reducible to propositions construed as sets of worlds or whatever. In Perry’s own writings this notion of meaning is measured in terms of behavioristically explanatory power, reminiscent of Stalnaker (1984; 1999).}

\footnote{This is the second place where our discussion crucially depends on beliefs being expressible in a language. As said before, when discussing the disquotational principle in 1.2.1, it may be a significant omission, but I will restrict attention to this class of beliefs.}

In Perry’s (1977; 1979; 1997) terminology: iff its direct expression in language (of thought) contains essential indexicals. \textit{Essential} here means exactly that the indexical cannot be replaced by a non-indexical description or other referring device, without changing the meaning of the belief expressed.

To avoid a terminological confusion sometimes found in the literature, note that we’re not talking about indexicals in reports; it is still a matter of debate whether essential indexicals in direct speech or in thought need to be reported with essential indexicals in a belief report’s complement. This confusion is perhaps further fed by the Latin (third person reflexive) \textit{se} which unfortunately seems to derive from the discussion about reports that ascribe a first person belief to a third person—a misnomer even for that class of
reports, which naturally should, and in most discussions does, include second and first person reports of first person beliefs or even, as I take the term, any reports of essentially indexical beliefs.

Consider:

(33) My abstract is awful

If Janell utters (33) she thereby expresses a de se belief, because the first person indexical possessive my is essential: for every description or (other) referential term \( \tau \) not containing any indexical, it is easy to construct a scenario where Janell does not believe herself to be \( \tau \), so that she would never believe or say \( \tau \)'s abstract is no good. So the indexical my in Janell's statement of her belief is essential. For example, take the proper name Janell for \( \tau \):

(34) Janell's abstract is awful

Assuming Janell has been struck with amnesia and doesn’t remember her own name suffices to show the non-equivalence of (34) and (33): not knowing her name, she might say (33), but she would not utter (34). In fact, for the first person case it doesn’t even help if we’d allow indexicals in our \( \tau \), so we could try things like:

(35) The abstract written by \( \{ \begin{align*} & \text{the person you are talking to} \\
& \text{the person I see in the mirror} \\
& \text{the person whose nose I’m now touching} \end{align*} \) is awful

I leave it to the reader to come up with examples about people who are mistaken about whose nose they’re touching, but it’s obvious how any such example can easily be adjusted so that Janell would be disposed to utter (33) but not the sentences in (35). So indeed, the belief Janell expressed with (33) is essentially indexical, or de se.

Note that the discussion of this example does not rely on reports. As I’ve said before, the terminology unfortunately derives from reports, i.e. sentences I would use to describe Janell’s belief to you (3rd person report, (36a)), or that someone would use to remind Janell of what is going on (2nd, (36b)), or that Janell herself would use, reflecting on what’s happening:

(36) a. Janell believes her abstract is awful
    b. You believe your abstract is awful
    c. I believe my abstract is awful
All of these are in this case reports of a first person *de se* belief, viz. the belief expressed originally with the essentially indexical \((33)\). Note also that, in the current terminology, \((36a)\) does not even contain an indexical; this *her* is an anaphoric pronoun.\(^{23}\)

So much then for first person *de se*. Other indexical beliefs include beliefs about the place where one is (expressed with *here*) or the present time (*now*). Proof: how to paraphrase *now* (together with the present tense in *is*) in “It is now 8:15PM” without smuggling in another essential indexical? The *name* of the time of *now*’s utterance won’t do: neither “8:15PM is 8:15PM” nor, if my watch were in fact 3 minutes fast, “8:12PM is 8:15PM” seems to carry the exact same truth conditions as the indexical belief. Other descriptions either rely on the indexicality of the present tense, or carry much more content than does *now* (e.g. *the time* Neon Genesis is supposed to start, according to the TV guide).

Just like with *I*, it is quite hard to come up with any synonymous term for *now*, even allowing indexicals to be used. Perhaps *the time at which this is said* would work (taking *this* to be self-referential)? This has in fact been proposed, by Reichenbach (1947), who takes it as a starting point for his *token-reflexive* analysis of the semantics of indexicals.\(^{24}\) We will ignore this account in favor of the more standard Kaplanian semantics to be discussed at length in 1.2.4.2.\(^{25}\) The important feature there is that all indexicals are

\(^{23}\)This is not your typical Binding Theoretic usage of the term *anaphor*. Based on (discourse) function, we classify person pronouns as follows: bound (or ‘donkey’) occurrences of *he, her* etc. are called *anaphoric*; referential *I, you, he [+pointing] etc. deictic* (or indexical); and *herself, yourself etc. reflexive* pronouns.

\(^{24}\)More recently, Perry (1997) successfully revived the token-reflexive analysis of indexicals attributed to Reichenbach (1947) and Burks (1949), by introducing a range of different levels of content/meaning, reminiscent of Kaplan’s (1989) split of Fregean *Sinn* into *Character* and *Content* (cf. 1.2.4.2), but more radical.

\(^{25}\)Apart from problems posed by the semantics of self-reference, the main problem I see with the token-reflexive account involves the standard argument for direct reference (=non-descriptivity) of indexicals. This argument, which originated with Kripke (1972) and was adapted for indexicals by Kaplan (1989), was meant to show that referential terms like indexicals and proper names (and perhaps others) cannot be paraphrased in descriptive terms. Applied to the Reichenbach analysis it runs as follows: Say, *now* means *the time of this utterance*. Then these terms should be substitutable for each other without changing the meanings of the utterances they occur in. But that’s false, because *Someone is speaking now* is true when someone says it, but only contingently: it is true or false in a possible world of evaluation depending on whether or not there are people speaking in that world on October 8, 2006, 14:44 CET, so for instance it’s false in a possible world where the entire human race was wiped out the day before (cf. the truth of the counterfactual *If the human race had been wiped out yesterday, nobody would be uttering anything right now*). But the Reichenbachian paraphrase *Someone is speaking at the time of this utterance*, with explicit reference to the speech act, becomes necessarily true.
interpreted in a manner that sets them apart from descriptive terms. His system achieves this by having two orthogonal dimensions of interpretation: contexts for indexicals, and indices for descriptive terms. Further, the class of indexicals is spanned by \textit{now}, \textit{actually} and \textit{I}, i.e. those three get a direct interpretation built into the model, while the meaning of other indexicals is defined in terms of them, but in such a way that the definition determines the complex indexical’s referent fully in the context dimension so as to avoid putting any descriptive meaning in the intensional part. In 1.B and 1.2.4.2 we see how this actually helps us get around the Kripke-Kaplan non-synonymy arguments, to which we now turn.

Taking a closer look at our definition of \textit{de se}, we may observe that the essentiality of indexicals in belief has actually been received wisdom in disguise since Kripke and Kaplan introduced their famous arguments establishing a fundamental difference between directly referential terms and their partly descriptive paraphrases in the 70’s.\textsuperscript{26} For example, \textit{you}, an indexical, seems to mean something like \textit{the addressee}, but actually they are not synonymous. We can tease apart the subtle semantic difference between indexicals and their descriptive counterparts by contrasting the sentences in (37):

\begin{enumerate}
\item You are the addressee
\item The addressee is the addressee
\item You are you
\end{enumerate}

My utterances of (37a) are usually true,\textsuperscript{27} but only contingently so, since instead of speaking to you I might have decided to talk to someone else: you are not \textit{necessarily} the addressee of my speech act. Substituting \textit{the addressee} for \textit{you} (37b), or the other way around (37c), may not change the actual truth value, but the meaning is not preserved, because both are now necessarily, and logically, true.

This argument is often reformulated by putting the appeal to necessity into the language, presumably to turn the necessity judgments into truth value judgments:

\begin{enumerate}
\item #You are necessarily the addressee
\item #?The addressee is necessarily the addressee
\item ?You are necessarily you
\end{enumerate}

\textsuperscript{26}Kripke’s (1972) examples primarily target the alleged synonymy between proper names (\textit{Aristotle}) and their proposed descriptive paraphrases (\textit{the Stagira born philosopher and teacher of Alexander the Great, who . . .}, or simply \textit{the individual called ‘Aristotle’}).

\textsuperscript{27}Unless there is no, or no single, identifiable, salient, addressee, in which case the utterance would suffer from presupposition failure and be uninterpretable rather than false.
However, scope ambiguities with respect to the overt operator of necessity actually cause these judgments to be indecisive (as indicated by the question marks in (38)). We return to these observations when discussing Kaplan (1989) and when setting up our own framework in chapter 3. For now, just remember how substitution can be used to bring out the semantic distinction between indexicals and descriptive terms.

Lewis and Perry merely reformulated this type of argument and called it the problem of the essential indexical. The prime exemplification of this problem is Perry’s (1977) Lingens, an amnesiac who doesn’t know where or who he is, but who, by reading an extremely detailed and up-to-date biography, comes to believe the proposition that Lingens is in the Stanford Library. We can imagine that this doesn’t help him if he still lacks the belief that he himself is Lingens. The missing belief is irreducibly indexical, expressible with I am Lingens and Perry’s point was to show the importance, ‘essentiality’, of indexicals in belief for explaining behavior. So it seems we are justified in coining a new term, de se, to refer to this important and sufficiently well-defined class of beliefs.

So far we have only considered what Perry (1997) termed automatic indexicals (e.g. I, now, yesterday,...), which automatically succeed in designating their actual referent, regardless of the speaker’s intention. Saying I, I always refer to me, Emar, even if I’m struck with amnesia or multiple personality syndrome, or say, think and have convinced you that I am Napoleon; the reference of an utterance of I is unaffected by the user’s mental state (as long as she can be said to speak English properly). Kaplan (1989) uses the term pure indexical for the same class, and opposes it with true demonstratives (e.g. this, there, that red car, you,...), indexicals that require something extra to secure successful reference. This extra is primarily the speaker’s intention—what does she want to refer to—which, for successful communication, can be clarified by, say, a nod of the head or even a genuine pointing; hence Perry’s term intentional indexicals. When combined with a properly recognized intention, these demonstratives are often no less essential then the automatic indexicals.

The uniform definition of de se in terms of essential indexicality notwithstanding, there are clear differences between pure and demonstrative indexicals, and many authors have (sometimes inconsistently and mostly implicitly) adopted a more restrictive notion of de se belief and especially de se reports, referring only to beliefs expressible essentially with pure indexicals. Witness for instance such definitions of de se belief and the de se/de re distinction

28He might for instance have consulted his Wikipedia entry at http://en.wikipedia.org/wiki/Rudolf_Lingens
1.2 Modes of attitude

...the access we have to ourselves is qualitatively different from the access (however direct) we have towards a certain ‘res’. [Chierchia 1989:3]

We shall call this subset of de se belief, those expressible only with automatic indexicals, pure de se belief, following a suggestion by von Stechow (1982). Other de se belief is often simply called de re. We shall stick to our original definition of de se however, calling those beliefs requiring intentional indexicals for their direct expression, intentional de se. It remains to be seen what will be the consequences of this way of dividing the cake for the de re/de se distinction, but first we need to focus on the de se/de dicto distinction.

1.2.3.2 De se vs. de dicto: properties vs. propositions

Now, on a more theoretical level, do we need special formal machinery to capture the semantics of the class of belief singled out above? Or can we perhaps reduce it to the familiar modal logic analysis of belief as a propositional attitude, that is, belief as a relation between a person and a set of worlds, as we did for de dicto (1.2.1), and ultimately, with the help of Kaplan’s (1969) acquaintance relations, with de re (1.2.2). The answer is almost unanimously no, propositions won’t do, and even singular propositions (propositions about objects, for instance structured propositions, cf. 1.2.2.1) won’t help. The most influential argument for this position was Lewis’ (1979a), which, moreover, came with a proper logic for the de se.29

29Lewis also added some thought examples of his own, the most famous being his “two gods” who are propositionally omniscient, i.e. they everything there is to know about the world they inhabit. The first, Zeus, lives on the highest mountain and occasionally throws down thunderbolts; the second, Jahweh, lives on the coldest mountain and throws down manna. And of course they each know that, that is, they know they’re in a world where Zeus lives on the highest and Jahweh on the coldest mountain. Despite all that, Lewis argues, it is still conceivable that they each not know who they are, i.e. that Zeus does not know whether he inhabits the coldest or the highest mountain. Lewis’ story can be seen as a more radical variation on Lingens’ predicament, for Lingens also knows all (relevant) propositions, but fails to realize who he is. Interestingly, there is a TV show called John Doe based on this very concept:

The series is about the life of John Doe, a mysterious man who rises from the primordial waters of an isolated island, possessing knowledge of literally everything in the world, yet having no memory of who—or even what—he is. [http://www.tv.com/john-doe/show/9049/summary.html]

(show canceled after one season in 2002)
Lewis’ argument starts with the above mentioned case of Lingens: With mere propositions, singular or general, there is no difference between what is expressed by Lingens uttering “Lingens is in Stanford” and “I am in Stanford”. Both express the proposition that this person, Lingens, has the property of being in Stanford. With structured propositions this is formalized as a pair consisting of the individual Lingens and the predicated property; in more traditional terms it is simply the set of worlds in which Lingens is in the extension of the predicate at that world:

\[
\begin{align*}
\{w \in W \mid \text{Lingens is in Stanford in } w\} \\
\text{[Lingens is in Stanford]}
\end{align*}
\]

The problem is that both utterances express propositions about Lingens, and that is ultimately because we agree with Kripke/Kaplan that it doesn’t make sense to interpret I descriptively, as the speaker, or Lingens as the so-and-so individual who is called ‘Lingens’—that would change the meaning: by using proper names or indexicals the speaker makes a claim about an individual directly, viz. about that individual she happens to refer to, whether she knows exactly who that is or not. On the propositional account of belief, then, this is the only plausible candidate for the object of the so-expressed belief of Lingens. From a given utterance of Lingens, we can, on the assumption that he’s sincere, conclude that he believes the proposition he expressed with that utterance. Ergo, both utterances express the same content, so the underlying belief is the same. But that can’t be right: because he’s lost, Lingens’ belief state only allows him to say “Lingens is in Stanford” (given his encyclopedic knowledge), obviously, the belief expressed with the indexical variant must be different. Next question: what is this difference in informational content, if not propositional?

Lewis’ answer is to shift from propositions to a slightly more fine-grained unit of information, properties, and, consequently, from belief (bel) as the primitive doxastic operator to self-ascription, and from worlds (in a Belief set) to contexts. Let me explain these moves. First of all, Lewis’ description of Lingens’ predicament in terms of properties and their self-ascription. Saying and believing: “Lingens is in Stanford” is Lingens’ way of locating himself in logical space, for which the traditional propositional account is suited perfectly (cf. the standard locution that propositions are ways of partitioning the space of possibilities, 1.1.1). In Lewis’ terms this is formulated—rather awkwardly, yet equivalently—as Lingens self-ascribing the property of inhabiting a possible world in which Lingens is indeed in Stanford. The propositionally equivalent eureka-moment “I am in Stanford”, however corresponds to Lin-
1.2 Modes of attitude
gens self-ascribing the property of being in Stanford. A different property altogether, one in which Lingens locates himself not (only) in logical space, but rather within a world, in real 3-dimensional space, viz. in the Stanford Library, or even in the body of one ‘Rudolf Lingens’. So, with self-ascription of properties we can do all that we can do with propositional belief, but we can also make some subtle distinctions that are actually needed.

Another example, also from Perry (1977): Heimson is mad. He thinks he is Hume. In terms of propositions, the belief he expresses when he says “I am Hume” is a contradiction,\(^{30}\) viz. the proposition that Heimson is Hume, so Heimson’s belief set is predicted to be empty too, i.e. he cannot distinguish any two worlds, he believes literally everything. This prediction is obviously too strong, moreover, in a way Heimson believes the same thing as Hume believed, viz. that they themselves are Hume, though, while each can confidently say “I am Hume”, the propositions so-expressed and therefore believed could not differ more: Hume expressed a tautology, Heimson a contradiction. In terms of properties this problem disappears completely: Heimson and Hume both self-ascribe the (non-trivial and non-empty) property of being Hume.

Now a bit more formal, what is a property, really? Basically, a property determines the extension of a predicate relative to a world of evaluation, i.e. it’s the intension of a predicate. In every possible world, a predicate like walks (walk) has an extension, the set of individuals that walk; the intension of walk, i.e. the property of walking, is the function from possible worlds to the predicate’s extension at that possible world. The set of properties is thus \((\wp D)^W\), the functions from worlds to sets of individuals. This set of course is in one-one correspondence with e.g. \((\wp W)^D\), and the form we will use from now on: \(\wp(W \times D)\), the set of \(\langle w, a \rangle\) pairs where \(a\) is an individual that has the given property in the world \(w\).\(^{31}\)

For example, recall lost Lingens and the proposition he believed on the classical account, (39). The properties he self-ascribes as evidenced by his utterances are:

\[(40)\]
- a. Lingens is in Stanford
  \(\leadsto \{\langle w, a \rangle \mid \text{Lingens is in Stanford in } w\}\)
- b. I am in Stanford
  \(\leadsto \{\langle w, a \rangle \mid a \text{ is in Stanford in } w\}\)

\(^{30}\)Remember, names just designate their actual bearers, so if the referents are distinct, the names are necessarily non-co-referential, so Heimson is Hume expresses a necessary falsehood equivalent to \(0=1\).

\(^{31}\)Trivial proof of the 2nd correspondence: map \(f \in (\wp D)^W\) to the set of \(\langle w, a \rangle\) pairs for which \(a \in f(w)\).
Chapter 1. Attitudes

One proposition, two distinct properties; self-ascribing (40a) is equivalent to believing a proposition (the $a$ is really a dummy there), but (40b) amounts to a real de se belief. The opposite—two propositions, one property—we find in the Heimson vs. Hume scenario that has them both self-ascribe the property below:

\[(41) \quad \text{I am Hume} \sim \{\langle w, a \rangle \mid a = \text{Hume (in } w)\}\]

With the move to properties and self-ascription, we announced a third Lewisian shift: from worlds to contexts. Lewis himself already notes that we can describe the self-ascribed properties equivalently as sets of centered worlds, or contexts. For this purpose, think of contexts as worlds from a particular point of view, a notion we may formalize as an ordered pair consisting of a world and its ‘center’, an individual.\(^{32}\) We need contexts like this anyway if we want to treat indexicality (=context-dependence) semantically.\(^{33}\) If we want to give a semantics of indexicals, we will have to say what is the semantic value (reference) of $I$, evaluated at a context. The answer is of course that the reference of $I$ at a context $\langle w, a \rangle$ is the center (agent, subject, speaker, point of view, . . .) $a$ of that context. A semantics that builds on this and does justice to indexicals’ direct reference is Kaplan’s (1989), detailed in 1.B. The main point for now is that contexts are worlds with extra structure, comprising at least a parameter specifying the center as we shall call it from now on. It’s easy to see that, formally, sets of contexts are the same as, or at least in one-to-one correspondence with, properties. The set of contexts in which “I am in Stanford” is true is the set of $\langle w, a \rangle$’s whose $a$ is in Stanford in $w$, and that’s exactly the property of being in Stanford as represented in (40b).

If we now turn to other indexicals we see that our properties are still not fine-grained enough, that we need more parameters in our contexts. Take now. An example belief de se about now is the one I express when I say “It’s now October 8, 2006”. What property do I self-ascribe? Obviously not being October 8, 2006, nor being (in existence) at October 8, 2006. What set of contexts make the sentence true? Thinking in terms of context it is obvious that for the interpretation of now we need an extra context parameter for

\(^{32}\)The center terminology, which I will use throughout, derives from centered world, which Lewis (1979a) ascribes to Quine (1969).

\(^{33}\)The same idea, extending possible worlds with other parameters of evaluation, is familiar from the semantics of modal logics that combine different modalities, such as tense and deontic modality, which necessitates the use of time/world pairs as evaluation points. Often these multi-parameterized intensional evaluation points are called indices. See 1.A.5 for tense logic and 1.B.2 for a full definition of contexts.
1.2 Modes of attitude

time. If contexts are triples containing a world, a person, and a time point, we can define the semantic value of now at a context \( \langle w, a, t \rangle \), as the context’s temporal center \( t \). Intuitively, the third context parameter is the time at which the center had the thought or utterance under consideration. Applied to the example:

(42)  It’s now October 8, 2006
\[ \sim \{ \langle w, a, t \rangle \mid t \in \text{October 8, 2006}, \text{in world } w \} \]

Not that this is no longer a property of individuals as such, but rather of individuals at specific times in their lives. Note that self-ascription is not an entirely accurate term anymore; in the above example I ascribe to myself at my ‘inner now’, or subjective temporal reference point, the property of being temporally located in October 8, 2006.

How about other indexicals, like here or that? One option is to add yet other contextual dimensions, like a place parameter representing the spatial location of the utterance situation (for here), and a parameter specifying the salient objects pointed to or intended by the subject (for that). Nothing stops us from going down this road, and indeed it has been done (Lewis 1972)—and criticized (Cresswell 1973). We will take an easier route, defended by Lewis (1981), and define the other dimensions of context in terms of the three we already have. The interpretation of here at context \( \langle w, a, t \rangle \) is the place where \( a \) is located at \( t \) in \( w \); and that refers to the object intended/pointed to by \( a \) at \( t \) in \( w \).

This is all I’ll give as argument, the unconvinced reader should check e.g. the systematic defense of this context reduction by Haas-Spohn (1994).

Thinking of Lewis’ analysis of de se as simply a shift from sets of possible worlds to sets of contexts as objects of attitude, we may try to retain the rest of the logic developed for propositional attitudes. By this I mean the definition of belief in terms of worlds and belief sets in 1.1.1 (details in 1.A), which we then later used as a semantics of de dicto belief in 1.2.1. What we

---

34 It would go too far to go into the treatment of the presupposed proximity contrast between this and that, or the treatment of extra descriptive material to form so-called complex demonstratives like that woman, a much debated subject. The issue is further complicated by the very common discourse deictic or anaphoric uses of these demonstrative forms.

35 Note that Reichenbach’s (1947) token-reflexive analysis discussed above can be construed as a 1-dimensional reduction of the context. In his account, all indexical parameters derive from the utterance. It suffices to assume that each utterance is a particular entity, occurs only in one world, has a unique speaker, and occurs at a single time interval. A similar 1-dimensional account is proposed by Lewis (1986:230), who admits world-bound time-slices of individuals to his ontology.
Chapter 1. Attitudes

did was equip every individual at a world with a belief set, Bel(w) ⊆ W—better yet, Bel(a, w, t) since different people have beliefs changing over time—defined as the set of worlds that that individual a cannot distinguish from (her conception of) the real world. Every proposition that is a superset of these so-called doxastic alternatives corresponds to a proposition that is, sloppily speaking, true in her representation of the world, i.e. she believes that proposition. To adapt that story to the property self-ascription account, just replace all worlds with contexts. Haas-Spohn (1994) provides an analogous test to the one cited on p. 9 to clarify what it means for a person to have a context as belief alternative: a context c is a belief alternative for a subject, if, were we to place that person in the world of that context, in the body/mind of the center of c, at the time of c, she would, upon careful investigation of her surroundings, agree that it is in fact the world as she knows it, from her own familiar point of view, at the time she thought was the time when the whole thing started. Summing up: with properties (43b) and an individual’s belief sets (43c) defined as sets of contexts (43a), the basic definition of doxastic logic, (10), p. 9, will be replaced by (43d):

(43) Let W be a set of worlds, D the domain of individuals, and T the set of times.
   a. contexts: c ∈ C ⊆ W × D × T
   b. properties: P ⊆ C
   c. belief sets: Bel : D × W × T → ℘C
   d. a self-ascribes property P at t in w iff Bel(a, w, t) ⊆ P

In conclusion, note that we have merely generalized the propositional account: contexts are more fine-grained objects than worlds, so from a set of self-ascribed contexts we can reconstruct the believed proposition as follows.

(44) a believes proposition p in w at t
    iff  a self-ascribes property \{⟨w', a', t'⟩ ∈ C | w' ∈ p⟩ at t in w

So, for completeness, assuming that de se belief complements are already given as properties, that mode can be cashed out as in (45b), and de dicto, (15), becomes (45b).36

36The real problems lie in the natural language semantics of reports: What utterances or report complements correspond to properties? How to compute what property is self-ascribed? For example, Colby believes that men are evil seems to ascribe to Colby a de dicto belief, whereas Colby believes that that man is evil may or may not report an essentially indexical belief of Colby’s, requiring, in the first case, the property of seeing/knowing a certain specific evil man. In other words, don’t confuse the operators believes\textsubscript{de se} or
1.2 Modes of attitude

(45) a. $a$ believes_{de se} $P$ at $t$ in $w$
    \[ \text{iff } Bel(a, w, t) \subseteq P \]

b. $a$ believes_{de dicto} $p$ at $t$ in $w$
    \[ \text{iff } Bel(a, w, t) \subseteq \{ (w', a', t') \in C \mid w' \in p \} \]

From now on, we will mostly use this heavier mechanism of contexts as modal parameters, since that allows a uniform treatment of _de se_ and _de dicto_. In the next section we'll ponder the consequences of this generalization for our account of _de re_ in 1.2.2. The main question there, leading up to the leading question of this work, will be: given the _de se_ inspired generalization to contexts for all belief, do we still need a distinction between _de se_ and _de re_?

1.2.3.3 _De se_ vs. _de re_: more terminological remarks

Above, we saw how essential indexicals in belief necessitate a move to properties. This should strengthen our case for choosing essential indexicality as such, rather than just the so-called _pure_ subclass, as the defining characteristic for the third mode of believing. But, as said before, many would classify the intentional _de se_ beliefs such as the one expressed with _that’s Ligia over there_ as _de re_, so let us now try and disentangle these notions properly.

Recall our definition of _de re_ in 1.2.2: A belief is _de re_ if it is about something external, which is only possible if the believer is sufficiently well acquainted with the object in question. The first thing to note is that _de se_ beliefs are always _de re_. This is because on our definition _de se_ belief requires that the believer can (or, is disposed to) state her belief with an indexical, and indexicals are _the_ way of picking out things in our immediate surroundings (context), with which we are ipso facto pretty well acquainted. We can restate the same argument starting out from Kaplan’s Principle 2: _Indexicals are directly referential_ (Kaplan 1989:492), which means they have no descriptive content but simply designate their actual referents from the context. Therefore, felicitous utterances of sentences containing essential indexicals express singular propositions (proposition about actual objects one is acquainted with), and finally, believing singular propositions amounts to _de re_ belief. Ergo: all _de se_ belief is _de re_ belief. As far as I know, this is almost universally recognized, although it is sometimes obscured by the tendency to shrink the application of the term _de re_ to refer only to the subset of _de re_ beliefs that are not _de se_. A more correct term for that complement class would be something like _mere de re_ (the deprecative _mere_ licensed by the fact that _de re_ is the more inclusive and therefore weaker notion).

\textit{believes}_{de dicto} with the English ‘believes’.
The obvious converse of this question has not gotten enough serious attention from philosophers, in my opinion: Is it perhaps not also the case that \textit{de se} subsumes \textit{de re}? Are there any mere \textit{de re} beliefs at all? In short, is there really a \textit{de re}/\textit{de se} distinction? Given our definitions it’s easy to see what would constitute a good answer: we have to provide a \textit{de re} non-\textit{de se} belief, or show that every \textit{de re} belief really does require indexicals if we want to state it directly. To answer this question, we must carefully reconsider the notions \textit{de se} and \textit{indexicality}. First, our liberality with respect to \textit{de se}, comprising pure and intentional \textit{de se}, may lead us to suspect that perhaps all \textit{de re} is \textit{de se}, contrary to received wisdom. This suspicion is further fed by the increasingly popular view that many referential terms are in fact \textit{hidden indexicals}, i.e. semantically they are context-dependent and directly referential in exactly the same way as \textit{you} and \textit{me}.

Take proper names. After Kaplan’s semantics of indexicals had become mainstream, it became almost trivial to take Kripke’s arguments and observations to show that proper names are indexicals, since they are prime examples of rigid designators whose reference is grounded in the actual context, i.e. in the baptism that actually took place in the utterance’s world. More needs to be (and has been) said, for instance about the additional role played by the referring intentions of the speaker in disambiguating in case of multiply borne names or incorrect usage, but treating names as (intentional) indexicals has proved a rather successful starting point for describing these issues in a formal semantic framework (see Burge 1973; Recanati 1993; Haas-Spohn 1994; Pelczar & Rainsbury 1998).

And there’s more: Haas-Spohn (1994) uses the arguments put forth by Kripke (1972), Putnam (1975) and Burge (1977), to argue for and outline a treatment of natural kind denoting terms (\textit{water, arthritis, \ldots}) as indexical. These treatments do require a gradual departure from the \textit{immediacy} of the context-dependence involved, we shift from mere visible surroundings and speaker intentions to dependence on intentions of the language community as a whole or even more remote concepts. In short, it has been convincingly argued that proper names and natural kind terms share the characteristic features of semantic indexicality with Kaplan’s traditional list of indexical expressions, viz. context-dependence (Principle 1 (Kaplan 1989:492)) and direct reference (or, rigidity: Principle 2). Are beliefs about water then per definition \textit{de se}? Are they even necessarily \textit{de re}? Here, terminology diverges, and I am not convinced for instance that the argument that all \textit{de se} is \textit{de re} (\textit{de se} = indexical reference, requires an intimate contextual acquaintance = \textit{de re}) above still goes through if we take hidden indexical \textit{de se} with its broad notion of context on board. If, however, we do decide that these are proper indexicals and therefore not counterexamples to the thesis that all
1.2 Modes of attitude

*de re* is *de se*, what possible *de re* non-*de se* candidates do we have left? What springs to mind are beliefs expressed with referentially used definite descriptions (Donnellan 1966). But the original discussion by Donnellan already shows a dependence on speaker’s intentions, which suggests that such descriptions as *the man holding a Martini glass* (referentially used) are also context-dependent yet rigid, i.e. indexical—perhaps semantically equivalent even to *complex demonstratives*, a subject briefly touched upon above in footnote 34 but beyond the scope of this work.

To conclude, it turns out to be remarkably hard to come up with a *de re* belief, i.e. about a thing with which you are properly acquainted, expressible in direct speech (or thought) without any kind of indexical (pure, intentional, or hidden). And with good reason, I believe, since that is exactly what indexicals are for: referring directly to objects in your immediate surroundings. However, I refrain from making any definite claims, and remain open for anyone’s suggestion of a proper counterexample. This is mainly because below I give a unified treatment of *de re* and *de se*, claiming that both traditional *de re* and traditional *de se*, however widely or narrowly construed, fall under a generalized notion of *de re*, independent of the outcome of the above debate, which is more of a terminological nature. Deviant terminology is thus the reason the thesis just defended—every *de re* belief is *de se*—may sound preposterous, since what philosophers often have in mind when talking about indexicals, are *pure indexicals*, and *de se* is often equated with pure, or even just first person, *de se* (Lewis himself being a case in point). From this point on, I will consistently use *pure indexicals* and *pure de se*, to refer to the smaller class, which indeed has some special properties when it comes to reporting, see 2.4.

1.2.4 Unifying *de re* and *de se*

1.2.4.1 Relational attitudes

Now, for a proper attempt at a unifying *de re* and *de se*, let us begin by simply generalizing the acquaintance-style analysis of *de re* (Kaplan 1969) to self-ascription of properties (Lewis 1979a), as proposed by Cresswell & von Stechow (1982) and von Stechow (1982). The idea is to make the general *de re* definition, repeated below, fit in with the property framework.

\[(46) \quad a \text{ believes}_{de\,re} \text{ of } b \text{ that it is } P \]
\[\text{iff} \quad \text{there is a concept } d \text{ s.t. } R(d,b,a) \text{ and } a \text{ believes}_{de\,re} \text{ of } b \text{ that it is } P \text{ under description } d\]

\[\text{[=}(31), \text{p. 33}]\]
We want to generalize the last part, which we had already reduced one step further, to:

\[(47) \quad \ldots \quad \text{and } a \text{ believes}_\text{de dicto} \text{ that } P(d) \quad \approx (31), \text{ p. 33}\]

Of course we could just fill in the definition of *de dicto* as a world-locating property, \((45b)\), without changing anything substantial to the Kaplanian story:

\[(48) \quad \ldots \quad \text{and } a \text{ self-ascribes the property of inhabiting a world where } d \text{ is } P\]

Cresswell & von Stechow take a different course; they allow for the possibility that the acquaintance relation underlying the description \(\alpha\) has essential indexical features, that is, the way of being acquainted with a *res* may lead the subject to have a *de re* belief that she herself could not express without indexicals. As argued above, this is quite often, if not universally, the case: almost all *de re* beliefs require indexical expressing and thus a real property analysis, not merely the *de dicto* border case. The question now is, how to modify \((47)\), constituting the crucial part of the definition of *de re* in \((46)\), so that it subsumes *de se*.

First of all, we take the indexicality of the acquaintance-based description into account, so we will replace the description \(\alpha\) with the acquaintance relation itself, relating the *res* to the ‘self’ at her ‘now’ in the complement of the belief. By so relating the *res* to the center, we essentially get a property-type complement: the set of contexts whose center is related to a *res* . . . (instead of *the set of worlds in which* \(\alpha\) . . .). Now, technically, \(\alpha\), meant to ‘represent’ an object descriptively, was a variable over individual concepts. In the same vein, acquaintance relations are really variables ranging over two-place predicates, i.e. over functions from worlds to pairs of individuals \(R\)-acquainted.

So, following Cresswell & von Stechow (1982), we replace Kaplan’s existential quantification over descriptive names \(d\) in \((46)\) with a quantification over acquaintance relations \(R\), and then the occurrence of \(d\) in the *de dicto* complement in \((47)\) becomes the *person or object that the center is \(R\)-acquainted with*. The self-ascribed property in full becomes the property of being \(R\)-related to a unique object that has \(P\).\(^{37}\)

\(^{37}\)The uniqueness is justified by the plausible assumption that suitable acquaintance relations can only relate you to one specific individual per world. In fact this uniqueness of the second argument, for all \(x\) there is exactly one \(y\) with \(R(x, y)\), is an important part of the definition of acquaintance relations. Without it, we would make wrong predictions: Tami is \(R\)-acquainted with Delbert, and then he says: “that Delbert guy (that I am \(R\)-
1.2 Modes of attitude

(49) \( a \) believes de re of \( b \) under acquaintance relation \( R \) that it has \( P \)
iff \( a \) self-ascribes \( \{ \langle w', a', t' \rangle \mid \text{the } b' \text{ with } R(a', b'), \text{ has } P, \text{ in } w' \text{ at } t' \} \)

Bringing it all together and spelling out the essential conditions making up \( R \) according to (30), we arrive at the following generalization of (31), to be known hereafter as the relational account of de re belief:

(50) \( a \) believes de re of \( b \) that it is \( P \)
iff there is a 2-place relation \( R \) s.t.
   (i) \( R \) is a suitably vivid acquaintance relation\(^{38}\)
   (ii) \( R(a, b) \)
   (iii) \( a \) self-ascribes the property of being \( R \)-related to a unique \( b' \) that is \( P \)

Let’s evaluate what we have here, first with Ortcutt. From the story on p. 16, and the reasoning in 1.2.2 it follows that Janell believes de re of Ortcutt that he is a genius under the acquaintance relation knowing . . . as a famous professor called ‘Ortcutt’. Proof: take \( R_1 \) to be that relation, i.e. \( R_1(x, y) = x \text{ read } y’s \text{ books and } y \text{ is a famous professor called ‘Ortcutt’} \). In the story it’s true that Janell self-ascribed the property of being acquainted with a unique \( y \) called ‘Ortcutt’ and famous, and that \( y \) being a genius, so clause (iii) in (50) is verified with \( R_1 \) as relation and being a genius as property. The other two clauses are also verified, as was shown in 1.2.2 already. On the other hand, Janell self-ascribes the property of being \( R_2 \)-acquainted with a loser, for \( R_2(x, y) = x \text{ reviewed } y’s \text{ abstract} \). If we assume, as before, that both \( R_1 \) and \( R_2 \) are suitably vivid, we get that Janell has the two de re beliefs Quine predicts, without having to assume a logical contradiction in her belief set.\(^{39}\)

\(^{38}\)In fact, because of uniqueness of the second argument, as described above in footnote 37, \( R \) can be viewed as a function, mapping an individual to the person she’s acquainted with at a certain time and world.

\(^{39}\)An interesting observation that will become important later, is that these verifying acquaintance relations, \( R_1 \) and \( R_2 \), are derived from the background story about Ortcutt, i.e. the (discourse) context in which the beliefs and their expression are situated. In fact, this is the crucial observation on which my own account of de re report semantics is built in chapter 3. Although this context dependence is a much neglected aspect of de re belief and reports, I am not the first to base a pragmatic account of attitudes or reports on it, cf. e.g. Aloni (2000), to be discussed in chapter 2.
Chapter 1. Attitudes

Thing is, with our enhanced definition of de re we can adequately capture (pure) de se belief as well. Take a traditional de se vs. de re scenario used to motivate the de se mode in 1.1.3.3: Janell has a de re belief expressed by abstract #037 should be accepted. Whether this is mere de re, hidden indexical, or intentional de se doesn’t matter now, but at least it’s not a pure de se. We model this by taking as acquaintance relation: \( R(x, y) = x \) reviewed \( y \) and \( y \) has number #037. The self-ascribed property is that of having reviewed an abstract numbered #037, which is so good it should be accepted. Then she learns it was actually her own abstract, that is, she adjusts her belief set so that only contexts remain whose center wrote abstract #037, leaving untouched the belief that that #037 should be accepted. It follows that then she self-ascribes the property of having written an abstract that should be accepted, viz. #037. That self-ascription obviously amounts to a pure de se belief, one she would express (when asked) with something like: “Remember that brilliant abstract we talked about, #037? Turns out it’s mine, so yes, I’m sure my abstract will be accepted,” featuring the egocentric indexical my (or something equivalent). This de se belief also falls under the generalized de re definition provided by the relational account, it’s just that the acquaintance relation under which the de re belief about the abstract is held has changed; the new acquaintance relation would be: \( R(x, y) = y \) is x’s abstract.

So we see again that pure de se falls under general relational belief. To be sure, consider the paradigmatic examples of Lingens and Heimson. First, Lingens has a de re belief about himself under the relation \( R(x, y) = x \) is reading \( y \)’s biography, with content \( P(x) = x \) is in Stanford. This changes to a pure de se belief about himself when he realizes I am Lingens. This belief corresponds to the set of contexts \( \{ (w, a, t) \mid a \) is Lingens\}, or, in other words, to a relational belief with content being Lingens under the acquaintance relation \( x = y \). The acquaintance relation of equality is needed here to fit this pure, egocentric de se belief into the de re format, where in general the self-ascribed property is of the form \( \{ (w, a, t) \mid \) in \( w \) at \( t \).Lingens = \( \eta [R(a, y)] \} \), given in (50). Combining these two de re beliefs (by intersection) gives Lingens’ final (pure de se) belief, that he is in Stanford. As for Heimson, there’s nothing new to be said, he believes something purely de se about himself, a belief he shares with Hume, viz. that he is Hume. According to the new de re definition this very belief is a de re belief of Heimson about Heimson himself, with content being Hume and acquaintance relation equality.

We conclude that, on the relational account, first person pure de se beliefs are simply de re beliefs under equality. This is exactly the result Lewis describes when he claimed de se falls under de re, a thesis ascribed to him, and elaborated upon by von Stechow (1982), based on the following passage:
1.2 Modes of attitude

Self-ascription of properties is ascription of properties to oneself under the relation of identity. Certainly identity is a relation of acquaintance par excellence. So belief *de se* falls under belief *de re*.

\[ \text{[Lewis 1979a:156]} \] 40

The same ideas apply to *de se* beliefs involving other context parameters. First let’s look at a temporal pure *de se* belief. The belief expressed by “It is now October 8, 2006” corresponds, as we saw, to a property of the form \{⟨w,a,t⟩ | t ∈ October 8, 2006\}. In relational terms, that should be a *de re* belief about the current time, under acquaintance the time currently experienced. In other words, take \( R \) such that at time \( t \), \( R(x,y) \) iff \( t = y \), then we get the right result if we fill it in in the self-ascribed property \{⟨w,a,t⟩ | in w at t:ry[R(a,y)] ∈ October 8, 2006\}. 41 The same goes for indexical beliefs featuring non-egocentric indexicals (that, there). To see that the belief expressed with “That man is crazy” is analyzable as *de re*, take \( R(x,y) \) to be \( x \) points at (or otherwise directs his audience’s attention to) \( y \) and \( y \) is a man.

It should be clear from these examples that the *de se* truth conditions, as modeled by property self-ascriptions, are properly captured by the enhanced definition of *de re* belief, as are classic *de re* cases like Ortcutt. In fact, we can precisely characterize the pure and other types of *de se* as subclasses of generalized *de re* by the type of acquaintance relation involved: pure *de se* for instance corresponds to *de re* belief under an egocentric acquaintance relation, i.e. relating \( x \) to \( y \) in \( w \) at \( t \) depending on certain non-intentional (‘automatic’) conditions on \( x, y, t \) and \( w \) (irreducibly depending on at least one of the context parameters \( x, t, w \) otherwise the property is just a proposition in disguise and the belief not purely *de se*). If, on the other hand \( x \)’s intention is relevant for determining whether \( R(x,y) \), the corresponding belief is either not indexical at all, or else then at least not pure, but hidden or intentional. So we arrive at the characterization of different subclasses of relational beliefs in figure 1.1, which sums up the above discussion. It further introduces some shorthands invented by von Stechow (1982) for referring to some common types of *de se* belief, like the first person *de se*, *de me*, and the intentional belief expressible with *that . . . , de hoc*. Conclusion: all classic examples of *de se* and *de re* are uniformly captured by a Kaplan (1969) inspired definition of

40In this passage, terminology deviates significantly, and tellingly, from ours: read first person *de se* belief, i.e. belief expressed with first person pronouns in direct speech for this occurrence of self-ascription of properties.

41A perhaps more insightful way to accomplish the same would be to extend the arity of acquaintance relations explicitly, i.e. \( R(w,t,a,y) \) as the relation of acquaintance holding between \( a \) and \( y \) in \( w \) at \( t \), so that we could take for \( R \) the relation relating every quadruple to its second component.
Chapter 1. Attitudes

Table 1.1: A classification of relational attitudes

<table>
<thead>
<tr>
<th>type</th>
<th>subtype</th>
<th>direct d.</th>
<th>in $w$ at $t$: $R(x, y)$ iff</th>
</tr>
</thead>
<tbody>
<tr>
<td>pure de se</td>
<td>$de$ me</td>
<td>I</td>
<td>$y = x$</td>
</tr>
<tr>
<td></td>
<td>$de$ nunc</td>
<td>now</td>
<td>$y = t$</td>
</tr>
<tr>
<td></td>
<td>$de$ hic</td>
<td>yesterday</td>
<td>$y = t - 1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>here</td>
<td>$y =$ the location of $x$ at $t$ in $w$</td>
</tr>
<tr>
<td>intentional de se</td>
<td>$de$ hoc</td>
<td>that</td>
<td>$x$ is pointing at $y$ in $w$ at $t$</td>
</tr>
<tr>
<td></td>
<td>$de$ hoc</td>
<td>that guy</td>
<td>$x$ is pointing at $y$ and $y$ is a guy in $w$ at $t$</td>
</tr>
<tr>
<td></td>
<td>$de$ te</td>
<td>you</td>
<td>$x$ is addressing $y$ in $w$ at $t$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SHE $+$ point</td>
<td>$x$ is pointing at $y$ and $y$ is female and not directly addressed in $w$ at $t$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>there</td>
<td>$x$ is pointing at a location $y$ not close to $x$ in $w$ at $t$</td>
</tr>
<tr>
<td>hidden de se or mere de re</td>
<td>Ligia</td>
<td>in $w$, $x$ is causally connected to a ‘baptism’ of $y$ with the name Ligia</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the man holding a Martini glass</td>
<td></td>
<td>$x$ intends to refer to $y$ by means of the man holding a Martini glass in $w$ at $t$</td>
</tr>
</tbody>
</table>

de re under a description/acquaintance relation, enhanced with the Lewisian notion of belief as property self-ascription. The idea of de se being but a subclass of de re in general can be traced back to Lewis (1979a), and was worked out in detail by von Stechow (1982).

Boër & Lycan (1980) actually proposed a stronger de se reductionist account, claiming that we don’t even need the second ingredient, the property ascription. They argue that de se is just traditional (say, structured proposition style) de re. The paradigm case of a first person de se belief (= de me) they analyze as follows:

\[ (51) \quad x \text{ believes } de \text{ me to be on fire iff } x \text{ believes}_{de \text{ re}} \text{ of } x \text{ that she’s on fire} \]

Such an account of course has trouble with the de se puzzles of Castañeda, Perry and Kaplan which were meant to show that $x$ can believe de re of $x$ that she’s on fire, without realizing it is her (in the mirror), and therefore not yet believing de se from the first person perspective that she’s on fire. Boër & Lycan’s (1980) ‘solution’ is not that interesting, they simply discard such data as mere pragmatic acceptability issues, maintaining there isn’t any semantic distinction. More importantly then, it raises the question how other Lewisian de se reductionists like von Stechow avoid this pitfall. In the version of his
account described in some detail above, the use of acquaintance relations and properties in (50) allows indexical descriptions/acquaintance relations, i.e. ones leading to an underlying de se belief. De se beliefs of every variety are easily definable, using the classification in figure 1.1. For example:

(52) a. $x$ believes de me to be on fire iff $x$ believes$_{de re}$ of $x$ that she’s on fire under the purely egocentric acquaintance relation equality

b. $x$ believes de nunc at $t_0$ that it is nice weather iff $x$ believes$_{de re}$ of $t_0$ that it is nice weather, under the purely egocentric acquaintance relation $R$, given by: in $w$ at $t$ $R(x, y)$ iff $y = t$

This concludes my exposition of a viable unifying account of de se and de re attitudes. The last part of this section explores a slightly different but very influential analysis of de se belief, in a framework based on a new semantic distinction: Kaplan’s two-dimensional semantics.

1.2.4.2 Kaplan: character vs. content

In this section we explore Kaplan’s (1989) Logic of Demonstratives, focusing on the theory of de re and de se attitudes it engenders. The motivation for this new logic however lies in the semantics of indexicals, so we’ll start out there.

In the first part of his paper, Kaplan sets out to defend and accommodate theoretically his aforementioned two principles governing the semantics of indexical expressions (p.492):

**Principle 1** The reference of indexicals depends on the context (in a systematic way)

**Principle 2** Indexicals are directly referential

Principle 1 is indeed obvious, for “if you and I both say ‘I’ we refer to different persons.” (p.492). Note that Kaplan’s notion of context is something like the (external) situation in which the utterance took place, i.e. the notion we have tried to capture formally as centered worlds, triples $(w, a, t)$. Although Kaplan is not so clear at this point, we can derive all needed dimensions of context, from speaker to intended demonstrations, from such context triples (cf. 1.2.3.1-1.2.3.2). The systematic way is the clearly statable descriptive meaning which lets us determine the referent of an indexical from its context of use, e.g. $I$ refers to the speaker of its utterance, or $that$ refers to the object pointed to. So much then for Principle 1, Principle 2 on the other hand does require some defense.
Principle 2 seems to deny indexicals all descriptive content other than their referent, which may come as a surprise, since we just saw that an indexical (unlike perhaps a proper name, the paradigm of direct reference) does have a clearly statable descriptive meaning. Kaplan maintains that this meaning, whatever its status, can not be part of the propositional/intensional content or what is said. To show this, Kaplan first considers embeddings under modals: in It is possible that tomorrow I’ll be here the indexicals I and here are unaffected by the operators that shift the possible world and timepoint of evaluation, they refer to the actual utterer and the location of her utterance respectively. If indexical content were part of the propositional content, we’d expect the possibility of a ‘narrow scope’ reading it is possible that tomorrow, a person will be speaking and that that person would then be at the place where she then is, but of course indexicals have no such readings. A prima facie alternative way to handle this would be to stipulate that indexicals always take widest scope with respect to any operator. This weaker consequence of Principle 2 is called the Fixity Thesis by Schlenker (2003), which may serve as a workable substitute for Principle 2 in most cases.42

However, as we have seen a couple times before,43 the Kripkean argument against descriptivity of names applies mutatis mutandi to indexicals: if I were not directly referential, its meaning (the speaker) would contribute to the propositional content, so I am speaking would mean the same thing as the speaker is speaking, but that’s nonsense since the first expresses a contingency, while the second is necessarily true (under one reading at least). The crucial point in this argument is that there are no operators, so scope can be of no help whatsoever. We are forced to accept Principle 2 (or the Fixity Thesis): indexicals are directly referential, in the sense that once the context has fixed a reference for the indexical, the descriptive meaning referred to in Principle 1 plays no role in determining the proposition expressed; it’s the individual actually picked out from the context that ‘goes into’ the proposition.44

The tension between the two principles, descriptive meaning governing systematic context-dependence vs. direct reference (no meaning, just refer-

42Perhaps this formulation is more palatable to those who dislike direct reference, for the Fixity Thesis allows one to cash out scope differently, say in terms of the independently motivated presupposition projection of definites. Such a maneuver will be explored further in chapter 3.


44 On Kaplan’s strict and literal account of direct reference, singular propositions are structured, they literally contain individuals as component parts. Formally, we will model direct reference in terms of rigidity in 1.B, as did Kaplan in his formal system. This means that instead of objects in propositions we represent these objects with a rigid designator that has no content but the object it refers to. Cf. 1.2.2.1.
ence), is resolved by acknowledging two kinds of meaning, called character and content, to replace the unitary Fregean notion of a Sinn or proposition. This distinction is intimately tied up with that between contexts of utterance and circumstances of evaluation, or context and world, for short. The two-dimensional semantics proposed by Kaplan goes something like this: A sentence or other expression in a language, only expresses (or contributes to) truth conditional content (= proposition = intension = what is said) if it is uttered in a certain context. Content is modeled as intensions, as in modal predicate logic. The content of my utterance of *I am stressed*, for example, is the set of worlds (or, in general, indices, if we take interpretive dimensions for tense into account, as we do in 1.B) in which I, that is, Emar, is stressed. This is the level of meaning in which indexicals are directly referential, the truth conditions depend on whether or not I am stressed in this or that world, not on other possible speakers and the fact that I am speaking is irrelevant as well. Accordingly, embedding under intensional operators, *Perhaps yesterday Colby believed that I was stressed*, doesn’t change this rigid reference of *I* to me, because, semantically, these operators operate at the level of content. Principle 1, then, is accounted for in the second level of meaning, character. The character of an expression is supposed to model the expression’s linguistic meaning, the rule that allows us to determine the content of an expression in each possible occasion of use, e.g. rules like *I* refers to the speaker of the context in which it was uttered. Summarizing: intensions are function from worlds to extensions; characters are functions from contexts to intensions.

For a term like *I*, the content in the current context where I’m writing this, is (rigidly) the extension, i.e. Emar, and the character is the function mapping each context to its center (speaker, writer, thinker). This character can be visualized in a diagram, plotting contexts on the y-axis, and worlds on the x-axis.

\[
\begin{array}{ccc}
\text{I} & w_0 & w_1 & w_2 \\
\hline
c_0 & \text{Emar} & \text{Emar} & \text{Emar} \\
c_1 & \text{Hume} & \text{Hume} & \text{Hume} \\
c_2 & \text{Lingens} & \text{Lingens} & \text{Lingens} \\
\end{array}
\]

where:
- \(c_0 = \langle w_0; \text{Emar}; \text{October 8, 2006} \rangle\), \(w_0\) is the actual world
- \(c_1 = \langle w_1; \text{Hume}; \text{May 31, 1750} \rangle\), \(w_1\) is very much like our actual world \(w_0\)
- \(c_2 = \langle w_2; \text{Lingens}; \text{October 8, 2006} \rangle\), \(w_2\) is the fictional world of the Heimson and Lingens thought experiments

Note that in such character diagrams, the cells are extensions, while the
intension of an expression in a context is easily read off from the row headed
by that context. In the case of a sentence, extensions are truth values,
intensions are functions from worlds to truth values, or equivalently sets of
worlds (classical propositions), and characters are functions from contexts to
sets of possible worlds, viz. those functions that, when applied to a context,
yield the proposition expressed in that context.

\[
\begin{array}{c|ccc}
  & w_0 & w_1 & w_2 \\
\hline
  c_0 & 0 & 0 & 1 \\
  c_1 & 0 & 0 & 0 \\
  c_2 & 1 & 0 & 1 \\
\end{array}
\]

added:
- in \(w_0\) Lingens is in Stanford
- in \(w_2\) I (Emar) and Lingens are both in Stanford

Kaplan argues that knowing a language implies, or rather consists in,
knowing the characters of its expressions (in addition to knowing its syntax,
and the modes of semantic composition). From this it does not follow that
everybody knows the content of what they or their friends are saying, for
often we don’t know exactly what context we are in. Lingens is an extreme
case in point, he doesn’t even know the content of his utterance of \(I \text{ or here}\),
i.e. he does not know what context he inhabits, one with Lingens as center
located in Stanford, or one with Dingens as center. This corresponds ex-
actly to the discussion of Lewis’ property theory, where we also said that a
person’s belief set contains various contexts, indistinguishable as far as the
believer is concerned, cf. 1.2.3.2. What is believed, or rather self-ascribed, on
that account were the properties described by supersets of these belief sets.
Kaplan has a different solution to the problems of essential indexicality. Ka-
plan agrees that Perry’s and Lewis’ examples convincingly show that belief
objects are not mere propositions. On his own theory that corresponds to the
fact that the Kaplanian content actually expressed may be something else
than expected when someone like Heimson sincerely says \(I \text{ am Hume}\), viz.
the (empty) set of worlds in which Heimson equals Hume. What Heimson
believes in the narrow, psychological sense—that which guides his actions—is
the character of his utterance, the function mapping each context onto the
proposition that the center of that context is Hume. In this sense, Heim-
son believes the same thing as Hume himself, the character of \(I \text{ am Hume}\)
even though the objective content expressed by their thoughts are completely
disjunct, a result paralleled by the property analysis.

Can Kaplan also handle the dual case, distinguishing pure \textit{de se} and
mere \textit{de re} as in the Lingens example, or Janell’s abstract? The answer
is yes, and in fact, Kaplan comes with his own example which has become paradigmatic of what is often (not always consistently, cf. p. 41) called the \textit{de re/de se} distinction: Imagine Kaplan looking in a mirror; he sees the person in the mirror’s pants are on fire, but, not realizing he’s looking at himself, he utters: “That guy’s pants are on fire!” The content he thereby expresses is that Kaplan’s pants are on fire, because the meaning/character of \textit{that guy} makes it refer to the object actually pointed at, i.e. Kaplan. The content is thus the same as when Kaplan 3 seconds later realizes it’s \textit{himself} he’s looking at, screaming: “Oh f***! It’s \textit{my} pants that are on fire”. The characters expressed are quite different however, because the characters of \textit{that guy(’s)} and \textit{my} are quite different, even though both are indexical, and both happen to refer (directly) to Kaplan in the two (rapidly succeeding) contexts considered. So it seems Kaplan’s characters come in handy for taking into account the indexical distinctions relevant to belief semantics.

Belief objects then are characters.\footnote{Or, perhaps closer to Kaplan’s way of speaking, belief objects are \textit{propositions under a certain character as mode of presentation}. But in the end such a proposal doesn’t really add anything since the proposition actually believed is easily reconstructed from the character by applying it to the actual context.} We would like to reduce this a little further, as done in classical doxastic logic with the definition in terms of belief sets or accessibility relations. This is not necessary however, nothing stops us from defining belief sets as simply sets of characters and leave it at that, so that \(x\) believes a character \(C\) iff \(C \in Bel^{\text{Char}}(x)\). But then, how about the observation above that (e.g. Lingens’) uncertainty can be modeled as not being able to distinguish among contexts, which is also the crucial observation in Lewisian property theory, leading to the idea that belief set is a useful notion that should be modeled as sets of contexts. Haas-Spohn’s (1994) test further fleshes out this notion of a context set, inspired among others by the work of Stalnaker (1984; 1999), which reinforces our idea that belief sets construed as sets of contexts are indeed relevant to a semantics of attitudes. The next section presents a way of translating the Kaplanian character theory to the context/belief set/property framework of 1.2.3.2.

### 1.2.4.3 Diagonals: from characters to properties

To synthesize the broadly Lewisian ideas about attitudes with Kaplanian characters, we must introduce a key notion in attitude semantics developed most extensively by Stalnaker (1978): \textit{diagonals}. A diagonal is an intension-like object collapsing the cognitively relevant information from a character into one dimension, taking only the information on the character’s top-left-to-bottom-right diagonal (in terms of character diagrams as exemplified in
(54)). Technically, the diagonal of a character maps each context \(c\) to the extension expressed in \(c\) and evaluated at the world of \(c\). Kaplan himself uses diagonals implicitly to define truth of a sentence in a context given the character: a sentence is true in a context iff it is actually true when uttered at that context (i.e. true in \(c\) at the world of \(c\)). The significance of the diagonal of a sentence character is therefore paraphrasable as the set of contexts in which the sentence could be uttered truly, just like a classical proposition was the set of worlds in which a sentence would be true (cf. 1.1.1), or rather, taking the 2-dimensional treatment of indexicality into account, the propositional content of a sentence uttered in a context is the set of counterfactual circumstances in which that utterance would be evaluated as true.

The diagonal unfortunately falls short of a reasonable explication of the linguistic meaning of a sentence, for knowing an expression’s meaning, especially as considered from a production perspective, consists not only in knowing in what contexts it would express something true, but also involves a knowledge of what would happen if the evaluation parameter were shifted to a counterfactual circumstance, by, say, a temporal operator.\(^{46}\) For cognitive significance, the other area where Kaplan wants to employ characters, it turns out that we can safely collapse the dependence on characters to their diagonals, a huge theoretical gain, mostly because it allows us to connect the Kaplanian theory with the property analysis.

To see why diagonals suffice for attitudes, first of all consider your favorite de se puzzle, say Lingens, as setup for a proof by example. Let’s add a context \(c_3\), one which Lingens cannot distinguish from his actual context \(c_2\), to the diagrams in (53) and (54) to make it more interesting:

\[
\begin{array}{l|ccc}
I \text{ am in Stanford} & w_0 & w_1 & w_2 \\
\hline
\text{c}_0 & 0 & 0 & 1 \\
\text{c}_1 & 0 & 0 & 0 \\
\text{c}_2 & 1 & 0 & 1 \\
\text{c}_3 & 0 & 0 & 0 \\
\end{array}
\]

added:

\(-\quad c_3 = \langle w_2; \text{Dingens; October 8, 2006}\rangle, \text{ and in } w_2 \text{ Dingens is at a library in Santa Cruz}\)

The thought experiment has it that Lingens doesn’t know who or where he is, even though he knows from his encyclopedic knowledge that he inhabits

\(^{46}\)In classical semantics such a question never arose, because these functions, meaning as truth conditions and meaning as behavior under intensional operators, are not and cannot be distinguished.
none other than $w_2$, in particular he knows that Lingens is in Stanford and Dingens in Santa Cruz. All this is captured by stipulating that his belief set contains (in a somewhat idealized case) two contexts $c_2$ and $c_3$ sharing $w_2$: $\text{Bel}(\text{Lingens}, t, w_2) = \{c_2, c_3\}$. Does Lingens believe what he would express with *I am in Stanford*? The traditional account would predict that he believes the proposition that he actually expressed, the third row in (55), $\{w_0, w_2\}$, because that is a superset of his classical belief set which would be $\{w_2\}$ (he knows which world he inhabits). But this is inadequate, for when asked “Where are you?”, he would never reply: “I am in Stanford.” On the diagonal version of Kaplanian character theory he would be predicted to answer thus iff the diagonal of the character in (55) were a superset of his enhanced (context) belief set. This is not the case since the diagonal contains only one context, $c_2$, in the other contexts, what he would have expressed would have been false, evaluated at the then actual world. In other words, the (characteristic function of the) diagonal is read off from the “coordinates” $(c_0, w_0)$, $(c_1, w_1)$, $(c_2, w_2)$, and $(c_3, w_2)$ in (55). It follows then that once he finds out he’s not Dingens, thereby discarding $c_3$ from his belief set, he is in a position to believe the character/sentence/diagonal of (55). To make the comparison with Lewisian self-ascription complete, note that the believed (or not) diagonals correspond exactly to the self-ascribed (or not) properties. For instance, the crucial property not self-ascribed by Lingens when he’s lost, describing his knowledge who and where he is, was the property of being Lingens in Stanford (cf. (40b), p. 43), formally \[ \{\langle w, a, t \rangle | a = \text{Lingens} \text{ and } a \text{ is, in } w \text{ at } t, \text{ in Stanford} \} \]. Restricting attention to the 4 contexts considered, this property is indeed the same as our diagonal: $\{c_2\}$.

What we showed here (admittedly, by way of an example) was that weakening Kaplan by looking at diagonals instead of whole characters, we get exactly the Lewisian property theory with its clean formalism, great explanatory power, and the link with classical notions like a person’s belief set. Note that this is not meant to suggest that Kaplan’s and Lewis’ frameworks are mutually reducible to each other, merely that the main theoretical assumptions and basic machinery are the same, and that this facilitates a decent comparison (cf. 1.2.4.4). In fact, skipping ahead somewhat to mention one tentative pro of Kaplanian character semantics, focus for example on the issue of compositionality with respect to the way we determine a self-ascribed property from an utterance or report (cf. discussion in and around footnote 36). Thorough discussion of compositionality however will have to wait until we’ve started on natural language semantics properly in chapter 2.

Finally, for those unconvinced by the example, let’s look at the general argument in favor of diagonals, in a formulation borrowed from von Stechow
& Zimmermann (2004:13-14), who in turn paraphrase the general ideas of Stalnaker (1978) as passed through by Zimmermann (1991) and Haas-Spohn (1994): Imagine characters $C$ and $C'$ with diagonals $\delta C$ and $\delta C'$, and imagine someone has a belief with cognitive significance $C$. This implies that there is a sentence $\sigma$ whose character is $C$ ([$\sigma$] = $C$) and that that person would be disposed to utter$^{47}$ when prompted to describe her belief.$^{48}$ From that disposition it follows that she considers herself to inhabit a context where that sentence, $\sigma$, would be uttered truly. This of course is exactly what’s captured by saying that her belief set is a subset of $\sigma$’s diagonal, i.e. $\delta C$. But then, by the same token she considers herself to inhabit a world where $\sigma'$ (a sentence with character $C'$ that she’d be disposed to utter as description of her belief that $C'$) expresses a truth. Finally, from uttering $\sigma'$ (or the disposition…) we conclude she believes [$\sigma'$], i.e. she believes $C'$. In less positive words, we’ve shown there are no two characterially distinct beliefs that coincide on the diagonal, so the cognitive significance of thought is justly reduced to diagonals. QED.

Let’s round off the discussion by summing up the advantages of restricting attention to the believed character’s diagonals: we get (i) the same representational power vis-a-vis purely indexical distinctions, (ii) an account of belief fully compatible with Lewisian self-ascribed properties, including (iii) compatibility with doxastic logic’s belief sets/accessibility given the refinement that belief alternatives are contexts rather than mere worlds.

### 1.2.4.4 Character theory vs. relational attitudes

In the remainder of this section we compare the two accounts of the full $de re$/$de se$ class of beliefs from table 1.1, to wit, Kaplan’s character theory with diagonals, and the relational account of $de re$/$de se$ developed in 1.2.4.1. Are there crucial differences, or can we use Kaplan as an unified analysis of $de re$/$de se$ on a par with the relational attitude account? In order to answer such a question we must straighten out some apparent differences in vocabulary.

Discussing $de re$ in 1.2.2, and continuing into the relational analysis in 1.2.4.1, we arrived at a notion of someone believing $de re$ some property of

---

$^{47}$Not “assent to” (von Stechow & Zimmermann 2004:13) since the intrinsic speaker change in assenting doesn’t preserve the perspective, i.e. the indexical mode of presentation formalized as character.

$^{48}$This is another case where we are forced to restrict ourselves to beliefs that can be put into language by the believer (cf. discussion about the disquotational principle on p. 22, and the definition of essential indexicals in 1.2.3.1, footnote 21). Again, I acknowledge that this constitutes a weakness in the argument, which may or may not be remedied, but one that will have to await another occasion.
something under some description. On the Kaplanian attitude semantics so far we only have belief as a 2-place relation, between a believer and a character (or in the simpler case, believer and diagonal). Interestingly, Kaplan himself speaks of believing a proposition under a mode of presentation given by the character. We've already noted however that in the Kaplanian framework the mode of presentation part, a character, is so rich that it renders the believed content part redundant, since we can decompose the simple a believes character C, in w at t into a believes proposition C(⟨w, a, t⟩) under mode of presentation C (or δC), in w at t.

So, in the character system, belief objects exhibit both aspects of the general Kaplanian theory of meaning: an objective, truth conditional (‘wide’) content, and a subjective, indexical (‘narrow’) character. Wide content, here, is the content actually expressed in virtue of the context you happen to be in. Recall that people are often confused about exactly which world or context they inhabit, so it often happens that we are unaware of, or mistaken about, which propositions are expressed by our utterances. Wide content is useful for giving actual realistic truth-conditions for sentences that express facts about the world, but cannot be used to explain behavior on the basis of our belief. The kind of content that’s ‘in the head’ and guides our actions, is called narrow content, and in Kaplan’s system this is successfully equated with characters. Applied to the Heimson vs. Hume case, the wide contents are the propositions that Heimson is Hume and the proposition that Hume is Hume: one false, one true, though Heimson is totally unaware of any such discrepancy and in fact thinks he expresses the second, true, proposition. The narrow contents are the same, viz. the character/diagonal of I am Hume, which accounts for the intuition that in this respect they share the same psychological state and are thus wont to act similarly (e.g. uttering “I wrote A Treatise of Human Nature” and preaching empiricism).

Turning to the relational account, we can see these same two aspects. In the central definition, (50) on p. 51, the first part—clauses (i), \( R \) is a vivid relation of acquaintance, and (ii), \( R(a, b) \) relates the believer \( a \) to an actual object \( b \) in the actual world. This corresponds, loosely speaking, to the belief’s wide content, since it is not required that the believer be aware what exactly this object \( b \) really is. For example, Janell believes de re of Ortcutt that he should be rejected, but she is totally unaware that she has such a belief about the individual she knows as ‘Professor Ortcutt’. The narrow part of the relational definition comes in clause (iii), the de dicto belief involving the mode of presentation rather than the actual individual. This de dicto belief, modeled as a property self-ascription, is what’s in the head. Moreover, as shown in the previous subsection, it actually coincides with the diagonal of the believed character on the Kaplanian analysis, so the
analogy makes perfect sense. To finish the example, in the Ortcutt case, this narrowly believed content is the property of being acquainted with someone who wrote an abstract labeled #042, and who is apparently some nitwit.

Conclusion: both unified analyses of de re/de se encode a wide and a narrow aspect of belief. The main difference lies in their accounts of the wide content: in the character theory it’s a full proposition, in the relational account it’s just the fact that a certain relation holds between believer and res in the real world. There is also a significant difference in the narrow parts, and that is the fact that believed character is a single homogeneous entity containing both (the subjective mode of presentation of) the res and the content predicated of it, while the relational account cleanly separates the (R-based description of) the res from the believed predicate P. One consequence of this will be that Kaplan’s analysis allows a nice compositional semantics of belief reports, but on the other hand it’ll be impossible to contextualize such a semantics, as we will do for the relational semantics, viz. by requiring that the acquaintance relation (which is the basis of the narrow description of the res) be determined from the context. These consequences for the semantics of reports will be discussed at length in the next chapter.

1.A Modal predicate logic

1.A.1 Syntax

A few preliminary words about notation: I use a courier-like font for expressions in the predicate logical object language, Greek letters as meta-variables over object language expressions (ξ for variables, α for constants, Π for predicates, ϕ for formulas), and calligraphic and italic symbols to denote set-theoretical entities in the meta-language.

(56)

a. Primitive symbols
   (i) Variables: x, y, x₁, x₂, ...
   (ii) Constants: a, b, a₁, a₂, ...
   (iii) For each n a set of n-place predicate symbols: walk, loser, love, =

b. Atomic formulas
   (i) If Π is an n-place predicate and τ₁, …, τₙ are terms (variables or constants), then Π(τ₁, …, τₙ) is a formula.

c. Complex formulas
   (i) If ϕ and ψ are formulas and ξ a variable, then ¬ϕ, ϕ ∧ ψ, □ϕ and ∃ξϕ are formulas.
1. A Modal predicate logic

1.A.2 Models

Model: $\mathcal{M} = \langle D, W, R, I \rangle$

- $D$ is the domain of individuals
- $W$ is the set of possible worlds
- $R \subseteq W \times W$ is an accessibility relation between worlds
- $I$ is the interpretation function:
  - (i) If $\alpha$ is a constant, $I(\alpha): W \rightarrow D$ (individual concept)
  - (ii) If $\Pi$ is an $n$-place predicate, $I(\Pi): W \rightarrow \wp(D^n)$ (property)

All semantic values below are to be thought of as relative to a model, but this will be left implicit in the notation, i.e. $\llbracket \varphi \rrbracket_w^{M, f}$ instead of $\llbracket \varphi \rrbracket_w^M$.

1.A.3 Free variables and assignments

- Atomic formulas
  - (i) $FV(\Pi(\tau_1, \ldots, \tau_n))$ is the subset of all variables among the $\tau_i$’s
- Complex formulas
  - (i) $FV(\varphi \land \psi) = FV(\varphi) \cup FV(\psi)$
  - (ii) $FV(\neg \varphi) = FV(\Box \varphi) = FV(\varphi)$
  - (iii) $FV(\exists \xi \varphi) = FV(\varphi) \setminus \{\xi\}$

A formula $\varphi$ with $FV(\varphi) = \emptyset$ is called a sentence of the language.

Formulas with free variables can only be interpreted with respect to a model and an assignment which maps at least the free variables to individuals. These will be called proper assignments for a formula:

- A partial function $f: \subseteq Var \rightarrow D$ is called a proper assignment for a formula $\varphi$ iff $FV(\varphi) \subseteq Dom(f)$

1.A.4 Semantics

A useful auxiliary notation:

$$g \succeq_A f \text{ iff } g \supseteq f \text{ and } Dom(g) \supseteq Dom(f) \cup A$$

Now, the full Tarskian, recursive definition of truth, or rather of semantic values. I assume that every formula below is interpreted w.r.t. a proper assignment for it, so we don’t have to worry about undefinedness of semantic values, for instance, the semantic value of a formula is always either 0 or 1.

- Semantic values of primitive symbols
Chapter 1. Attitudes

(61)  a. $\mathcal{M} \models_w \varphi$ iff $[\varphi]_w^f = 1$

b. $\mathcal{N} \models \varphi$ iff $[\varphi]_w^f = W$

c. $\models \varphi$ iff for all models $\mathcal{M}$: $\mathcal{M} \models \varphi$

d. $\varphi \models \psi$ iff for all models $\mathcal{M}$: if $\mathcal{M} \models \varphi$ then $\mathcal{M} \models \psi$

And then some final notation we will occasionally use:

(63)  a. Syntax: if $\epsilon$ is an intension-denoting expression, then $^\forall \epsilon$ is an
1.B Kaplan’s 2-dimensional semantics

expression that denotes the corresponding extension

b. Semantics: $[\epsilon]^f_w = [\epsilon]^f(w)$

1.A.5 Extra modalities: belief and tense

(64) a. Syntax: If $\tau$ is a term and $\varphi$ a formula, $\text{Bel}_{\tau}\varphi$ is a formula
b. Model: $\text{Bel}: D \times W \rightarrow \varphi(W)$
c. Semantics: $[\text{Bel}_{\tau}\varphi]^f_w = 1$ iff $[\varphi]^f_w \supseteq \text{Bel}([\tau]^f_w, w)$

(65) a. Syntax: If $\varphi$ is a formula, then so are:
   (i) $p\varphi$ (sometime in the past, $\varphi$ was the case)
   (ii) $f\varphi$ (sometime in the future, $\varphi$)
   (iii) $h\varphi = \neg p\neg \varphi$ ($\varphi$ has always been the case)
   (iv) $g\varphi = \neg f\neg \varphi$ ($\varphi$ is always going to be the case)

   b. Model:
      (i) $T$ is a set of times
      (ii) $\prec \subseteq T \times T$ denotes temporal precedence
      (iii) $\langle T, \prec \rangle$ is a linear ordering

   c. Semantics:
      (i) $[p\varphi]^f_t = 1$ iff there exists a $t' < t$ in $T$ with $[\varphi]^f_{t'} = 1$
      (ii) $[f\varphi]^f_t = 1$ iff there exists a $t' > t$ in $T$ with $[\varphi]^f_{t'} = 1$

1.B Kaplan’s 2-dimensional semantics

Kaplan’s (1989) ‘Logic of Demonstratives’ is an extension of the semantics of modal predicate logic. It essentially adds a new parameter for evaluation, the context, which is kept separate from the modal/tense dimension called circumstances of evaluation, or simply indices. The main motivation comes from the semantics of indexicals ($I$, here, now, tomorrow) and demonstratives ($this$, that book). A correct semantics of indexicals, he argues, requires meaning to be split up into truth conditional/referential content and descriptive/linguistic meaning. Kaplan also offers an indirect speech operator SAY, which has since then been adapted in the literature to an analysis of attitude reports.

1.B.1 Syntax

The syntax is that of predicate logic with metaphysical modality $\Box$, $\Diamond$, tense operators $P$, $F$, $H$, $G$, (we will add belief $\text{Bel}$ below) as discussed in 1.A. In addition, Kaplan adds a few special, indexical constants, and operators:
Chapter 1. Attitudes

1.B.2 2D models

(67) Model: \( \mathcal{M} = \langle D, T, <, W, R, \Pi \rangle \)

a. Contexts: \( \mathcal{C} = W \times D \times T \ni c = \langle w_c, a_c, t_c \rangle \)

b. Indices: \( \mathcal{I} = W \times T \ni i = \langle w_i, t_i \rangle \)

1.B.3 Semantics

The definitions of free variables, assignment functions, and proper assignments for a given formula are the same as before. The truth definition is an extension of (61) and it runs like this (again, everything relative to proper assignments):

(68)

a. Semantic values of primitive symbols
   
   (i) \( \llbracket \xi \rrbracket^f_{c,i} = \llbracket \xi \rrbracket^f = f(\xi) \in D \)
   
   (ii) \( \llbracket \alpha \rrbracket^f_{c,i} = I(\alpha) : \mathcal{C} \rightarrow \mathcal{I} \rightarrow D \), \( \llbracket \alpha \rrbracket^f_{c,i} = \llbracket \alpha \rrbracket^f_{c,i} (i) = \llbracket \alpha \rrbracket^f (c)(i) \)
   
   (iii) \( \llbracket \Pi \rrbracket^f_{c,i} = I(\Pi) : \mathcal{C} \rightarrow \mathcal{I} \rightarrow \mathcal{P}(D^n) \), \( \llbracket \Pi \rrbracket^f_{c,i} = \llbracket \Pi \rrbracket^f_{c,i} (i) = \llbracket \Pi \rrbracket^f (c)(i) \)

b. Semantic values of atomic formula
   
   (i) \( \llbracket \Pi(\tau_1, \ldots, \tau_n) \rrbracket^f_{c,i} = 1 \text{ iff } \langle \llbracket \tau_1 \rrbracket^f_{c,i}, \ldots, \llbracket \tau_n \rrbracket^f_{c,i} \rangle \in \llbracket \Pi \rrbracket^f_{c,i} \)
   
   (ii) \( \llbracket p \phi \rrbracket^f_{c,i} = 1 \text{ iff there is a } t < t_c \text{ s.t. } \llbracket \phi \rrbracket^f_{c,i} = 1 \)
   
   (iii) \( \llbracket y \phi \rrbracket^f_{c,i} = 1 \text{ iff for } t, \text{ the day before } t_c : \llbracket \phi \rrbracket^f_{c,i} = 1 \)

The definitions of proper assignments and proper assignments for sentences are as before.
The basic semantic unit, \([\ldots]\), defined in (68) is called the character of a term or formula. It is a function from contexts to ordinary one-dimensional intensions (i.e. individual concepts or propositions). Only in a context, does a sentence have real truth conditional content (intension): \([\varphi]^{c}\), the proposition expressed by \(\varphi\) in \(c\).

The system also allows for finer distinctions regarding notions like truth, truth in a model, and logical truth:

\[
\begin{align*}
\text{(69)} & \quad a. \ M \models^{f,c} \varphi \text{ iff } [\varphi]^{f,c}_{(w_{c},t_{c})} = 1 \\
& \quad b. \ M \models^{f,c} \varphi \text{ iff } [\varphi]^{f,c}_{(w_{c},t_{c})} = 1 \\
& \quad c. \ M \models^{f} \varphi \text{ iff for all } c: M \models^{f,c} \varphi \\
& \quad d. \ M \models^{f} \varphi \text{ iff for all models } M: M \models^{f} \varphi \\
& \quad e. \ M \models^{f} \psi \text{ iff for all models } M \text{ and all } c: \text{ if } M \models^{f,c} \varphi \text{ then } M \models^{f,c} \psi
\end{align*}
\]

The logic is usually enriched with operators, for rigidification and diagonalization:

\[
\begin{align*}
\text{(70)} & \quad a. \ \text{Rigidification: } [\text{DTHAT } \epsilon]^{f,c}_{(w_{c},t_{c})} = [\epsilon]^{f,c}_{(w_{c},t_{c})} \\
& \quad b. \ \text{Diagonal: } [\varphi]^{f} = \{ c \in C \mid [\varphi]^{f,c}_{(w_{c},t_{c})} = 1 \}
\end{align*}
\]

A purely semantic diagonal operator, working on character directly, will be useful too:

\[
\text{(71)} \quad \text{For every character } C \in \{0,1\}^{TC}: \delta C = \{ c \in C \mid C(c)((w_{c},t_{c})) = 1 \}
\]

\[
\text{Note: } [\varphi]^{f} = \delta [\varphi]^{f}
\]

1.B.4 Adding ‘says’ and ‘believes’

Believing or saying something relates an individual to a character:

\[
\text{(72)} \quad \text{Direct speech and (cognitive significance of) belief}
\]

\[
\begin{align*}
& \quad a. \ \text{Syntax: If } \tau \text{ is a term and } \varphi \text{ a formula, } \text{SAY}_{\tau} \varphi \text{ and } \text{BEL}_{\tau} \varphi \text{ are formulas} \\
& \quad b. \ \text{Model: } \text{Say, Bel}^{\text{Char}}: D \times I \to \varphi[C \to \{I \to \{0,1\}] \\
& \quad c. \ \text{Semantics: } [\text{BEL}_{\tau} \varphi]^{f,c}_{i} = 1 \text{ iff } [\varphi]^{f} \in \text{Bel}^{\text{Char}}([\tau]^{f,c}_{i}, i) \text{ (idem for SAY)}
\end{align*}
\]

When it comes to report semantics, Kaplan focuses more on indirect speech than on belief reports, but since we are currently interested in the latter, we follow the natural extensions to Kaplan’s notoriously sketchy analysis of attitudes and reports.
Chapter 1. Attitudes

(73) a. Model: $Bel^C : D \times I \rightarrow \wp(C)$

b. Semantics: $\llbracket Bel_{\tau} \varphi \rrbracket_f^i = 1$ iff $\\varphi \\models_f \supseteq Bel^C(\llbracket \tau \rrbracket_f^i, i)$

To make the formulations in (72) and (73) compatible, we relate the definitions of $Bel^C$ and $Bel_{Char}$ as follows:

(74) For every character $C$, $a \in D$ and $i \in I$: $C \in Bel_{Char}(a, i)$ iff $\delta C \supseteq Bel^C(a, i)$

Kaplan’s semantics of reports is discussed at length in 2.3.4, but for completeness, here’s the central definitions:

(75) Indirect speech and belief reports

a. Syntax: If $\tau$ is a term and $\varphi$ a formula, $Say^{\text{that}}_{\tau} \varphi$ and $Bel^{\text{that}}_{\tau} \varphi$ are formulas

b. Semantics:

(i) $\llbracket Say^{\text{that}}_{\tau} \varphi \rrbracket_f^i = 1$ iff there is a character $C \in Say(\llbracket \tau \rrbracket_f^i, i)$ s.t. $\llbracket \varphi \rrbracket_f^i = C(\langle w_i, \llbracket \tau \rrbracket_f^i, t_i \rangle)$

(ii) $\llbracket Bel^{\text{that}}_{\tau} \varphi \rrbracket_f^i = 1$ iff there is a character $C$ s.t. $\delta C \supseteq Bel^C(\llbracket \tau \rrbracket_f^i, i)$ and $\llbracket \varphi \rrbracket_f^i = C(\langle \llbracket \tau \rrbracket_f^i, t_i, w_i \rangle)$

---

49 This requires that $Bel_{Char}$ be closed under logical consequence as defined (indirectly, only for expressible characters) in (69e). This diagonal closure condition is also part and parcel of the reconstruction of ‘Adding “Says”’ by von Stechow (2001:3) and von Stechow & Zimmermann (2004:8), who credit Haas-Spohn (1994) with the idea that believing a character is self-ascription of its diagonal, from which the closure immediately follows. This particular closure presumably does not hold of speech reporting, because arguably, a diagonal carries the cognitive significance, but not the full linguistic meaning of a sentence.
Chapter 2

Attitude reports

In this chapter we shift the focus from logico-philosophical explorations to their application in natural language semantics.

2.1 *De dicto, de re, and de se reports?*

The leading question is in what way the intuitive threefold distinction in attitude modes that we started out with in 1.1.3 applies to attitude reports. As has been stressed above, attitudes are mental states while attitude reports are natural language sentences. To be precise, we will consider sentences of the form \( NP \text{ believes (that) } S \) meant to convey that the subject bears the attitude of belief towards the proposition expressed by the complement sentence. As we saw (p. 14), it is a matter of some debate whether a theory of attitudes should yields a theory of the truth conditions of their reports, or vice versa, but in any case, we must take care not two confuse these two distinct fields of study.

Nevertheless, given the traditional threefold distinction\(^1\) in modes of attitude we may follow common terminological practice and posit a parallel subcategorization of the linguistic class of attitude reports, i.e. a distinction in mode of reporting, based on whether the ascribed attitude is *de se, de re* or *de dicto*. A *de re* report, for example, will be defined as a sentence that ascribes someone a *de re* belief. Of course, a first complication is the fact that report classes so-defined will inherit the (essentially terminological) difficulties we already encountered in dividing up the class of attitudes in this way, moreover, the theoretical reductions and unifications we were eventually led to (e.g. the reduction of *de re* to *de dicto* in 1.2.2.5, and the unification

\(^1\)Note the *traditional* here, since as I argued in chapter 1, a consistent rethinking of the definitions of *de re* and *de se* may reduce these two to a single class.
of de re and de se as relational attitudes in 1.2.4.1), will automatically carry
over to reports. Apart from those preliminary complications, the interesting
question is, does a de dicto/de re/de se division along these lines correspond
to any interesting, independent linguistic categorization?

At first sight, there indeed appear to be some straightforward correspon-
dences between the sentence structure of the report and the mode of the
reported belief. For instance, quantifiers like every and a, or bare plurals
(1a) generally correlate with ascriptions of de dicto beliefs, referential ex-
pressions like proper names (1b) and indexicals with de re (i.e. anything
ranging from mere de re, to intentional and even pure de se), and finally
subjectless infinitival complements (1c) with pure de se:

(1) a. Ellsworth Kimmel believes pretty girls always win
    b. Ellsworth believes of Arata Suggs that she’ll win
    c. Ellsworth believes to be on the winning side

These examples exemplify unambiguous de dicto, de re and de se reports
respectively, though admittedly (1b) (and for some even (1c)) is not very
natural English. The sentence in (1a) is true iff Ellsworth Kimmel would
assent to something like “Pretty girls always win”; (1b) iff she knows this
person, Arata Suggs (under whatever guise), and thinks of that person that
she is bound to win; finally, (1c) iff she could express her belief with a pure
indexical, first person subject: “I am winning.” The truth conditions of these
three examples thus support the above mentioned heuristics for recovering
report mode from surface clues. Our goal now is to investigate whether it is
possible to turn these rough heuristics into a purely linguistic categorization
of belief reports, so that the mode of the reported belief can be derived
directly, preferably compositionally, from the reporting sentence’s surface
structure.

Note however that many reports are ambiguous\(^2\) with respect to the at-
titude mode, (2) for example is ambiguous between de dicto and de re:

(2) Ellsworth believes the prettiest girl will win

The de dicto reading is the one that reports an attitude Ellsworth has about
very pretty people in general. The de dicto truth conditions thus resemble
those of (1a). But there is also a different reading, de re, on which (2)
reports an attitude Ellsworth has towards a certain specific individual she
is acquainted with. If we assume that this person is in fact Arata, then
the truth conditions of the de re reading correspond to those of (1b). The

\(^2\)Let’s settle on the term ambiguous for now, later we will make this more precise and
discuss whether it’s a case of syntactic ambiguity or rather pragmatic underspecification.
2.1 De dicto, de re, and de se reports?

The crucial difference with the *de dicto* reading is that it need not be part of the content of Ellsworth’s thought that this person is pretty, it may simply be the reporter’s own preferred way of picking out Arata. Therefore, only on the *de re* reading does the report remain true under (some) substitutions of co-referential terms such as in this scenario *Arata*, or *That girl over there [pointing at Arata]* for the prettiest girl. On the other hand, the *de dicto* reading can be true even if Ellsworth is not in any way acquainted with Arata, who is in fact the prettiest girl. I will consider the *de dicto/de re* ambiguity from a more formal semantic perspective in 2.2, where, in particular, we take a closer look at a scopal account of this ambiguity.

Another infamous family of ambiguous reports is exemplified by (3), which has been argued to be ambiguous between a *de re* and *de se* reading:

(3) Ellsworth believes she’s going to win

Taking the pronoun to be anaphoric (i.e. not deictic) and referring back to Ellsworth, we can presumably read this sentence *de re* about Ellsworth, i.e. true iff Ellsworth believes *de re* of Ellsworth that she’s going to win. On this reading (3) would be true if Ellsworth were to point at a TV screen, saying: “Whoa, that girl is going to win” provided she’s actually pointing at her own image on TV, but irrespective of whether or not she recognizes herself. The pure *de se* reading on the other hand is a report of the first person attitude she would verbalize as “I’m going to win.” We take up the *de re/de se* distinction, and especially the contrast between (3) and (1c), in 2.3.

Let us summarize our initial observations on modes of attitude reporting by bringing together the 5 sentences discussed above: (1a), (1b), (1c), (2), and (3). To show that the observed linguistic differences in reporting modes are real, semantic differences, let’s assume that actually, Arata is just a nickname for Ellsworth that she is not aware of, and that Ellsworth is really the prettiest. Such a setup might lead an unsuspecting reader to believe that given these additional assumptions surely all five must have the same truth conditions. A closer look reveals that she’d be wrong, for consider the following situations:

A. Ellsworth believes that the prettiest always win, it’s just part of the system in her eyes. She does not believe that she is the prettiest, so she’s a little disappointed. She herself describes her belief as: “Unfortunately, the prettiest always win, so I have no chance.”

Scene A is obviously one where Ellsworth’s belief is *de dicto*. And I’m sure
your intuitions will agree that, if I have to choose one of the 5 reports above, I best choose (1a) to describe this situation, though (1b) would be true too.

**B.** Now, Ellsworth is not convinced that being pretty is any guarantee for winning. Instead, at a party, drunk, hearing people talk about Arata Suggs while watching Ms. Suggs speech on TV, she says: “That girl, Arata, that everyone’s talking about, she’s really smart. I’m sure she’ll win” In fact this Arata Suggs is none other than herself, she just didn’t recognize the image on the TV, nor did she know of this nickname the others use to refer to her.

In B, Ellsworth arguably has a *de re* belief, for she’s quite vividly acquainted with someone referred to as Arata Suggs, through watching the TV broadcast and overhearing the talk of the people in the room. Assuming that’s enough to establish a proper acquaintance relation, we should expect that (1b) is the optimal way of reporting on the goings on in B, and I believe it is indeed by far the best (of the five possibilities given). It’s equally clear that (1a) and (1c) are completely out, but intuitions are somewhat muddy with the the ambiguous forms (2) and (3). Certainly these are suboptimal, but if we assume that it is salient common knowledge among speaker and intended audience (though Ellsworth herself remains oblivious) that in fact Ellsworth Kimmel = the prettiest = Arata Suggs, we can perhaps assign a true reading to these as well.

**C.** As in B, Ellsworth does not think looks matter, but she is very confident and would tell anyone: “I am going to win.”

Finally, in C the belief is purely *de se*, *de me* to be precise, and anybody will agree that both (3) and (1c) would describe this situation. And again, after a bit of philosophical exercise, and stressing that the audience and the speaker have just been informed about the various guises of Ms. Kimmel, I trust you’ll be able to squeeze out the true readings of (2) and (1b).

To summarize these basic intuitions for future reference, table 2.1 contains the basic truth value judgments that will keep us busy for most of this chapter. We can conclude somewhat preliminarily that in English, the *de dicto - de re - (pure) de se* division of attitude modes is not straightforwardly grammaticalized in the language of attitude reports. In other words, a belief’s linguistic realization in a report is often un(der)specified with respect to mode of ascribed attitude. This is not to say that every report can report any mode of attitude: on the edges of the gray areas we can single out some unambiguous prototypical examples, as witness (1). It remains to be seen what kind of ambiguity or underspecification we are dealing with here.
2.1 De dicto and de re reports

A. B. C.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1a)</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(2)</td>
<td>1</td>
<td>?1</td>
<td>?1</td>
</tr>
<tr>
<td>(1b)</td>
<td>0</td>
<td>1</td>
<td>?1</td>
</tr>
<tr>
<td>(3)</td>
<td>0</td>
<td>?1</td>
<td>1</td>
</tr>
<tr>
<td>(1c)</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.1: Intuitions

Is it perhaps just another case of context-dependence? If so, what kind? And how to account for its contextual resolution in a formal semantic framework? These are among the main themes of the current work.

2.2 De dicto and de re reports

In this section we take a look at the de dicto/de re distinction for reports. We will use predicate logical formulas, as introduced in chapter 1, to represent adequate truth conditions (logical forms, LF’s) for belief reports that ascribe de dicto or de re attitudes. An important theme is that, as we just saw in 2.1, many sentences admit both de dicto and de re interpretations. The traditional view has it that this should be thought of as a scope ambiguity at the level of logical form, i.e. a syntax/semantics ambiguity. In 2.2.1 we elaborate on this traditional account a little, ending with a brief note on a more recent, dynamic, pragmasemantic reincarnation of the scope idea. It must be noted that mere scopal accounts (syntactic or pragmatic) fail to take into account the modes of presentation for de re readings, which, as the Orcutt example (pp. 16-17, 27ff) convinced us, are essential ingredients of de re believing. In 2.2.2 I implement the Kaplanian solution with acquaintance relations in a report semantics, fully in accordance with the corresponding belief account of 1.2.2.

2.2.1 Scope ambiguity

Our goal is a semantics for belief reports, i.e. we need a way of mapping sentences of English to their truth conditions. For now, we look at sentences of the form NP believes that S and represent truth conditions as formulas in modal predicate logic with belief (cf. 1.A). In this section we want to be able to correctly predict the judgments of (1a), (1b) and (2) as plotted out in table 2.1.

A simplified translation of (1a) into modal predicate logic would look
something like (4) (ignoring all manner of sophistication with respect to bare plurals and adverbial quantification):

(4) Ellsworth Kimmel believes pretty girls always win

\[ \triangleright \text{BEL}_a \forall x [\text{pretty}(x) \rightarrow \text{win}(x)] \]

This representation squares with the account of *de dicto* belief as propositional belief, cf. 1.2.1.

For (1b) we can posit a wide scope representation to give the *de re* truth conditions. This is in accordance with 1.2.2.1, where scoping the *res* outside of an intensional operator was shown to be one of the possibilities for representing *singular propositions*, the traditional way of modeling *de re* belief.\(^3\)

(5) Ellsworth believes of Arata Suggs that she’ll win

\[ \triangleright \exists x [x = a \land \text{BEL}_a \text{win}(x)] \]

According to the traditional scope account, sentences like (2), with a definite description as embedded subject, have two logical forms: one where the description takes narrow scope, the other wide: \(^4\)

(6) Ellsworth believes the prettiest girl will win

a. \[ \triangleright \text{BEL}_a \text{win}(\#y \text{prettiest}(y)) \]

b. \[ \triangleright \exists x [x = \#y \text{prettiest}(y) \land \text{BEL}_a \text{win}(x)] \]

The truth conditions are thus reasonably well taken care of, i.e. as well as singular proposition can capture *de re* belief—which is not at all perfect, cf. 1.2.2.3 and 2.2.2. In fact, the above distribution of logical forms does correctly capture our semantic judgments, as seen by comparing our predictions in table 2.2 with the original intuitions of table 2.1.

The debate can now concentrate on what kind of embedded NPs should give rise to such a narrow/wide ambiguity, e.g. are there wide (*de re*) readings for embedded indefinites? *de dicto* readings for proper names? Answers to such questions will be relevant to the matter of *compositionality* for these

---

\(^3\) The logically equivalent alternative puts a as the argument of win in the scope of BEL, while assuming that constants like a are rigid designators that therefore in a sense take wide-scope semantically, though not at the representational level (cf. 1.2.2.1): \(\text{BEL}_a \text{win}(a)\)

\(^4\) A truly Russellian treatment of definite descriptions would claim that (proto-)representations like in (6) in fact already abbreviate the two readings (narrow and wide) intended here. This is because, on such an account, a formula with a \(\tau\)-term is really an abbreviation that can to be expanded to a real predicate logical formula in different ways. The definition of \(\tau\)-terms we use here makes them of the same type as individual constants: \(\llbracket \tau x . \text{win}(x) \rrbracket_w = \text{the winner in } w\). Cf. 1.A.
2.2 De dicto and de re reports

<table>
<thead>
<tr>
<th></th>
<th>If</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1a)</td>
<td>(4)</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(1b)</td>
<td>(5)</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(2)</td>
<td>(6a)+(6b)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.2: Predictions of the scopal account of de dicto/de re

sentences, e.g. can we build these representations with a word-by-word translation mechanism? Note in this connection that we are essentially postulating a syntactic ambiguity here: on one derivation the (potential) res is interpreted in situ (de dicto), on the other it’s moved outside the attitude verb, leaving a singular proposition as belief object (de re). With the advent of dynamic, or rather, underspecified, semantics it has become possible to abandon the syntactic ambiguity in favor of a pragmatic one. In such a theory, only a single, underspecified LF is generated, and it is left to contextual, pragmatic factors to determine whether we end up with wide or narrow scope truth conditions. We will study this and other applications of dynamic semantics, underspecification and context-dependence in detail in chapter 3.

2.2.2 Kaplan (1969) revisited: vivid names and de re as de dicto

Unfortunately, scopal accounts in themselves are still deficient with respect to the logical form of the de re readings, as shown by Kaplan’s (1969) Orttcutt example. In 1.2.2 we reconstructed Kaplan’s argument and saw that simple singular propositions will not do. Even for a de re belief the mode of presentation, a description of the res, should be taken into account in the belief complement. Kaplan (1969), as we saw, took vivid names relating the believer to her res to provide such descriptions in such a way that the direct, wide scope, object dependence is actually removed from the complements of de re beliefs. In this sense he reduced de re belief to a subclass of de dicto belief, viz. those where the description of the subject of the belief is given by a believer’s vivid name for an actual res. In this section we incorporate these insights of Kaplan (1969) into the semantics of de re reports.

The Kaplanian definition of de re belief can be straightforwardly turned into a semantics of de re reports (and de re readings of ambiguous reports), as in (7). The $d$ is, again, a variable over descriptions of individuals (i.e. individual concepts), and $R$ is a fixed abbreviation (its interpretation should be part of the model) of the fact that, in this case, $d$ is a vivid name, of $b$ and for $a$. A new shorthand is ‘NP$_a$’, which is, roughly, a placeholder for an
NP referring to the object \( a \) (and logically represented as \( a \)), in other words:

\[
\llbracket \text{NP}_a \rrbracket_w = \llbracket a \rrbracket_w = a.
\]

(7) \[ \llbracket \text{NP}_a \text{ believes of NP}_b \text{ that it is VP}_P \rrbracket_w = 1 \quad \text{iff} \quad \text{there is an individual concept } d \text{ s.t. } R(d,b,a) \text{ and } a \text{ believes}_{\text{de dicto}} \text{ that whatever is } d \text{ is } P \]

\[ \supset \exists d[R(d,b,a) \land \text{BEL}_a P(\nu d)] \]

On the basis of this semantics, Kaplan would assign the following If to our \textit{de re} report (1b):

(8) Ellsworth believes of Arata that she will win

\[ \supset \exists d[R(d,e,a) \land \text{BEL}_a \text{win}(\nu d)] \]

If we assume that the \textit{de dicto} If’s of (1a) and (2) remain the same as before, the \textit{de dicto}/\textit{de re} ambiguity takes on the following form:

(9) Ellsworth believes the prettiest girl will win

\begin{align*}
\text{a.} & \quad \text{BEL}_a \text{win}(\nu \text{prettiest}(y)) \\
\text{b.} & \quad \exists d[R(d,e,\nu x. \text{prettiest}(x)) \land \text{BEL}_a \text{win}(\nu d)]
\end{align*}

What, semantically speaking, are the predictions of this account with respect to the pattern described by table 2.1? The logical form of the \textit{de dicto} sentence (1a) and of the \textit{de dicto} reading of (2) remain untouched, but the \textit{de re} logical forms are as in (9b), i.e. true if there is some suitable, vivid name \( d \) under which Ellsworth holds the belief about herself (=Arata=the prettiest). For (8) in B we could take the girl seen on TV that everyone talks about, named ‘Arata Suggs’ as \( d \); in C we could take myself, Ellsworth Kimmel (or a more elaborate description picking out Ellsworth for Ellsworth). And of course, for the \textit{de re} reading (9b), we could take the same \( d \). One problem is that, arguably, seeing a person on TV giving a speech while being drunk does not quite qualify as constituting a vivid name. However, as discussed in 1.2.2, we assume that indirect perception can count as \textit{de re} and take it from there. Summing up, the predictions of this ‘singular-proposition-less’ Kaplanian alternative to the scopal account discussed above, then completely match those of the classical account, as shown in table (9).

Note also that with this machinery we can do justice to reports in Ortcutt scenarios like 1.1.3.2’s Janell who believes of Ortcutt, under different guises, that he is brilliant and that he’s an embarrassment. Both sentences \textit{Janell believes Ortcutt is brilliant} and \textit{Janell believes Ortcutt is dumb} can be true at the same time because we’re dealing with distinct vivid names of Ortcutt for Janell. Also, the report \textit{Janell believes Ortcutt is brilliant and dumb} is
predicted to be false (unless Janell is logically insane) for there is no vivid name that would verify it, cf. 1.2.2.

There have been substantial refinements and fundamental criticisms of the Kaplanian account of *de re* belief. One such refinement is the incorporation of the context-dependence of vivid names, which is shown to have some advantages over the current existential quantification. The hardest critique has to do with the alleged non-compositionality of the approach. Both matters have direct parallels in the more general relational account that can be seen as an extension of the vivid names analysis covering the whole class of *de re* and *de se* belief reports. Therefore, we will take up the issues of context-dependence and compositionality after introducing the relational semantics, in 2.3.2 and 2.3.3, respectively.

### 2.3 *De se* and *de re*

This section extends the two unifying accounts of *de re/de se* belief described in 1.2.3, in order to provide a working semantics of belief reports. These are, respectively, the semantics of relational attitudes (Lewis 1979a; von Stechow 1982) developed for belief in 1.2.4.1; and Kaplan’s (1989) theory of belief as a characterial operator developed in 1.2.4.2 and 1.8. For both resulting types of reductionist semantic frameworks, we consider some pros and cons, and possible amendments. Section 2.4 will then discuss some recent, data-driven criticism against the unification of *de re* and (pure) *de se* as such. My aim in the end (chapter 3) is to revive the reductionist thesis—that *de re* and *de se* reports form one homogeneous class—by using a novel pragmatic-semantic framework that allows a treatment of the problematic scenarios and counterexamples that we will encounter in the current chapter.

#### 2.3.1 The semantics of relational attitude

In 2.2.2 we turned Kaplan’s (1969) analysis of *de re* belief into an analysis of *de re* belief reports. Synthesizing Kaplan’s vivid names and Lewis’ properties, we got what we termed the *relational* theory of *de re* and *de se* belief.

<table>
<thead>
<tr>
<th></th>
<th>If</th>
<th>A.</th>
<th>B.</th>
<th>C.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1a)</td>
<td>(4)</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(1b)</td>
<td>(8)</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(2)</td>
<td>(9a)+(9b)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.3: Predictions of Kaplan’s vivid name analysis
Now, we’ll look at the predictions of that theory, when construed as a theory of belief reports, as was first done by Cresswell & von Stechow (1982) in Cresswell’s (1973) framework of \( \lambda \)-categorial LFs including structured propositions. The theory has later been modified and applied to a wide range of de re phenomena such as sequence of tense (Abusch 1997) and ellipsis of belief reports (Reinhart 1990).

Postponing everything related to compositional derivation of LF (c.q. \( \text{lf} \)) to 2.3.3, the basic definition in relational belief semantics is the following adaptation of chapter 1’s (50):

\[
\begin{align*}
(10) & \quad [\text{NP}_a \ \text{believes of} \ \text{NP}_b \ \text{that it is} \ \text{VP}_P]_{w} = 1 \\
& \text{iff} \quad \text{there is a 2-place relation} \ R \ \text{s.t.} \\
& \quad (i) \quad R \ \text{is a suitably vivid acquaintance relation} \\
& \quad (ii) \quad R(a, b) \\
& \quad (iii) \quad a \ \text{self-ascribes the property of being} \ R\text{-related to a unique} \\
& \quad b' \ \text{that is} \ P \quad [\approx (50), \ p. 51]
\end{align*}
\]

As with the vivid name semantics we want to capture these truth conditions in a formula. With respect to the third clause that means we need a way to talk about non-atomic properties. For this purpose we can use the well-known apparatus of \textit{lambda abstraction}, introduced properly in appendix 2.B. With the help of a \( \lambda \), a new property-taking BEL-operator called \( \text{BEL}^* \) (denoting self-ascription), and leaving clause (10i) implicit, a relational \( \text{lf} \) looks as follows:

\[
\begin{align*}
(11) & \quad \text{NP}_a \ \text{believes of} \ \text{NP}_b \ \text{that it is} \ \text{VP}_P \\
& \quad \gg \quad \exists R[R(a, b) \land \text{BEL}^*_a \lambda x[P(\iota y[R(x, y)])]]
\end{align*}
\]

For the \textit{de re} readings considered in the previous section nothing changes compared to the vivid name truth conditions of 2.2.2. For completeness’ sake:

\[
\begin{align*}
(12) & \quad \text{Ellsworth believes of Arata that she’ll win} \\
& \quad \gg \quad \exists R[R(e, a) \land \text{BEL}^*_e \lambda x[\text{win}(\iota y[R(x, y)])]] \quad [= (1b)] \\
(13) & \quad \text{Ellsworth believes the prettiest girl will win} \\
& \quad a. \quad \text{BEL}^*_e \lambda x[\text{win}(\iota y[\text{prettiest}(x)])] \quad [= (9a)] \\
& \quad b. \quad \exists R[R(e, x[\text{prettiest}(x)]) \land \text{BEL}^*_e \lambda x[\text{win}(\iota y[R(x, y)])]] \quad [= (9a)]
\end{align*}
\]

For current purposes, the \textit{de re} \( \text{lf} \)s in (12) and (13b) are fully equivalent to those in (8) and (9), and in fact have the same truth values in A-C, since \( a \) and \( x[\text{prettiest}(x)] \) are co-referential. For both we can take the exact
same verifying acquaintance relations in B and C as we did on p. 79, viz. in B, $R(x, y)$ holds iff $x$ sees $y$ on TV and $y$ is called ‘Arata Suggs’, and for C, $R(x, y)$ iff $x = y$. We’ve left the de dicto readings as they were, as witness (13a), so that in fact all our predictions so far match those of the previous section.

The point of the Lewisian extension however was to account for the somewhat subtler de re/de se distinction, in particular the difference between scenarios B and C, and consequently between their observationally most optimal reports (1b) and (1c), and then the ambiguous (3). So let us apply the relational analysis to these examples. As stated above, (1b) is true in both B and C with $R$’s as described earlier. But what is the logical form of the infinitive construction (1c), and how do we interpret the anaphoric pronoun in (3)?

We start with the latter, (3), since it’s trivial on the current account: the pronoun is simply a bound variable, i.e. bound by the antecedent Ellsworth. To represent this in predicate logic we’ll just copy the constant $e$:

(14) Ellsworth believes she’s going to win

$\exists R[R(e, e) \land \text{BEL}^* \lambda x[\text{win}(\text{if}\{R(x, y)\})]]$

Note that the pronoun that seems to be in the syntactic scope of believes, has wide scope in the logical form, i.e. outside the intensional $\text{BEL}^*$ operator, at the same level as its antecedent, Ellsworth. The fact that the syntactically embedded subject has wide scope at If, is one of the most striking features of the relational analysis and its Kaplanian (1969) predecessor, and has been a source of considerable and ongoing debate. On the one hand, syntacticians have argued that such ‘long movement’ is totally ad hoc and ‘unprecedented’, and semanticists have complained about, or obscured, the obstacles such movement causes for compositional semantics. We discuss this matter further in 2.3.3. But before that, and throughout the rest of this work, I will point out the advantages of the wide scope account as such over the various in situ analyses, like the quantified character analysis of 2.3.4, or the anti-reductionist systems of 2.4.

The first such advantage we encounter right here, with (3)’s If (14): this single logical form accounts for what are sometimes referred to as the sentence’s de re and de se readings, meaning that it’s true in both B and C. This is easily verified by taking the same $R$’s we’ve been using above, $x$ sees

---

5We could have added a quantified variable $x$ and set $x = e$, or instead, represent proper names like Ellsworth as properties (as in some versions of DRT) or as quantifiers (as in Montague grammar). In fact our simplification only works because all $e$’s occur unembedded, otherwise we’d have to assume at least rigidity of these constants.
Chapter 2. Attitude reports

...TV and y is called ‘Arata Suggs’ for B, and x=y for C. A big advantage is that we need not postulate any ambiguity, the relational $f$ existentially quantifies over the acquaintance relation under which the belief holds. The $f$ is therefore compatible with any kind of relational attitude summed up in table 1.1 p. 54, to wit, pure de se (as in C), intentional de se (as in B), hidden de se, or even mere de re.

The unification of de re and de se described in 1.2.4.1 and exploited here in the semantic representation of (3) is thus not only a nice and sound theoretical reduction unifying two modes of attitudes, it also helps to get the correct predictions for sentences like (3) in an elegant way that avoids ambiguity or special treatments of embedded pronouns. As noted, compositionality remains an issue, and we will return to it shortly, but not before we take a look at the flip side of de re/de se unification: if the logical form of a report like (3) is compatible with both pure and other de re and de se scenarios, how do we account for pure de se reports like the infinitival one in (1c)?

This was the puzzle posed first by Chierchia (1989) who takes the observed contrast between (3) and (1c) as a starting point to argue against a unified treatment of pure and other de re/de se belief reports. Let’s dwell on those judgments a bit, since they’re crucial, and these particular examples may not be the most convincing. First of all, this may be due to the fact that the English believes to be construction is not unanimously accepted as grammatical, though it is attested on the web (of course): The author believes to be aware of related intellectual property rights [. . .] or He believes to have found dung and paw prints that resemble a human double fist.

Fortunately, even if you don’t want to accept these sentences as grammatical, note that the same acceptability contrast can be observed with other attitude verbs in English, and, secondly, with belief reports in other languages. An example of the first type is the following hope report:

(15) Essie Beard is to appear on a game show on TV. After the make-up artists are through with her, and being quite stressed out, she doesn’t even recognize herself when she sees herself in a mirror. Thinking she’s seeing her opponent, she mutters: “I hope you lose.”

---

6From www.ietf.org/ietf/IPR/infineon_ietf_ipr.pdf Note: I have been unsuccessful in filtering out examples where believes to be/have has come about as the result of raising plus extraction, as in e.g. the stuff that he believes to be true. Since examples of this type are ubiquitous on the web, I can’t give any estimate of the relative frequency of the construction we’re interested in.

7http://home.fuse.net/rschaffner/panther.htm

8In the literature, many resort to hopes to + infinitive as the paradigm example of a pure de se report, cf. e.g. (Schlenker 2003; Percus & Sauerland 2003a).
a. Essie hopes that she’ll lose.
b. #Essie hopes to lose.

In the direct expression of her hope Essie uses an irreducible, intentional indexical pronoun you, so this scenario exemplifies an intentional but not pure de se hope of Essie. The hope is also de re in the general sense, viz. de re about Essie herself, for that is the referent of you. To get a true reading of (15a) we need to bear in mind that it is a report made by an informed speaker (say, me), to an audience that shares the knowledge about Essie’s peculiar predicament of not recognizing her own mirror image. In such a setup it is conceivable that I felicitously utter (15a), to highlight the irony of the situation. Verbalizing some of the above meta-remarks in the actual discourse surrounding the report will make the contrast even clearer:

(16) So, apparently, Essie hopes that she’ll lose, though she doesn’t realize that herself.

While such an addition may or may not increase the acceptability of the that + third person pronoun report, it definitely decreases the acceptability of the infinitival report, which becomes truly incoherent/ungrammatical, regardless of further context:

(17) *So, apparently, Essie hopes to lose, though she doesn’t realize that herself.

Unfortunately, the semantics of the attitude hope poses a couple of independent technical difficulties—having to do with the asymmetric dependence of hope on belief discussed on p. 34—which is why I want to stick with belief, to keep things manageable when getting our hands dirty formalizing in chapter 3.

Another way out is to admit that, perhaps believes to be is not perfect English, but then view it as a word-by-word translation gloss of a similar report in one of the languages where these constructions are undeniably unmarked. For simplicity I stick with my (Pseudo-)English example, but as a matter of fact nothing hinges on English, as is recognized especially by linguists. In fact, Chierchia’s original examples were in Italian where crede di essere [believes to be] is the standard way of ascribing pure de se beliefs,

---

9This is a bit sloppy. One true direct discourse version would be for instance the imperative “Lose!” with implicit deixis to the addressee. Meta-note: reporting such imperatives while retaining the imperative mood (in what might be called an ‘imperative report’) happens to raise a related issue, viz. that of 2nd person pure de se reports, cf. footnotes 34 and 119.
and I can confirm that in Dutch the analogous constructions *denkt te* and *meent te* + infinitive are OK, as corroborated by a quick glance at the first page of (the more than 10,000) Google hits on "*denkt te*".\(^\text{10}\)

So let us conclude that there are infinitival reports that are incompatible with *de re* attitudes about oneself that are not pure *de se* (e.g. mirror or TV scenarios like above), but true if the belief is pure *de se*. To be a bit more precise, read *1st person de se* for *pure de se*. In fact, let us agree to focus on the domain of person from here on, since it makes for clear examples and e.g. tense is a much more complicated subject, where many independent difficulties (even fundamental questions like, is tense anaphoric or deictic?) would distract us from the *de re/de se* issues that concern us here. Before continuing Chierchia’s argument against the kind of unified treatment provided by the relational analysis, first we need some syntactic theory.

Since the relational analysis can only be applied to sentences that give us two individuals: a believer, \(a\) (the denotation of the matrix subject), and a \(res, b\) (denotation of embedded subject).\(^\text{11}\) The infinitival complement in (1c) however has no subject, at least not on the surface. However, it has been argued on syntactic grounds that such clauses have to have a grammatical subject. Following common syntactic practice, we call the invisible subject of an embedded infinitival or gerundial clause \(PRO\):

\[
\text{(18) Ellsworth believes PRO to be on the winning side} \quad \approx \text{(1c)}
\]

It seems reasonable to assume, if we want to apply the relational analysis, that this PRO refers to Ellsworth (e), so we’d get the following logical form:

\[
\text{(19) } \exists R[R(e,e)] \land \text{bel}^* \lambda x \left[ \text{win}(\forall y[R(x,y)]) \right]
\]

This, unfortunately, is the logical form of an unspecified *de re/de se* belief, exactly what we got in (14). In other words, we have failed to do justice to the fact that language can specify, to a certain degree, what type of *de se* attitude is reported. This argument led Chierchia to conclude that infinitival (and gerundial) attitude complements have different logical forms, one not captured by the relational account, but compatible only with first person attitudes (or 2nd for command reports, cf. footnotes 34 and 35, p. 119). We will detail Chierchia’s “*de se separatist*” counterproposal, and the positive

---

\(^\text{10}\)Luckily, Dutch does not allow the raising construction (*she believes him to be crazy*) that complicated the Google search for English examples, cf. footnote 6. In particular, counting Ghits gives: #*hij denkt te*: #*hij denkt dat* hij \(\approx 1\):3

\(^\text{11}\)Easily generalized to reports about \(n\) \(res\): put them all in a single \(n\)-tuple. Cf. 2.3.3.
2.3 De se and de re

arguments for it, in 2.4 below, but first we’ll delve a bit deeper into the relational account and discuss some other arguable weaknesses and ways in which to fix those.

\[
\begin{align*}
(1a) & & (13a) & & 1 & & 0 & & 0 \\
(1b) & & (12) & & 0 & & 1 & & 1 \\
(2) & & (13a)+(13b) & & 1 & & 1 & & 1 \\
(3) & & (14) & & 0 & & 1 & & 1 \\
(1c) & & (19) & & 0 & & 1 & & 1
\end{align*}
\]

Table 2.4: Predictions of the relational analysis of de re/de se

2.3.2 The pragmatics of acquaintance

As an alternative to de re analyses in terms of unrestricted quantification over concepts (cf. discussion on pp. 30-31), Aloni (2000) proposes a pragmatic analysis where the res description is to be determined in the context. To this end the essentially Kaplanian definition, (7), p. 78, is modified to select vivid names from a context set \( C \) that represents salient information from the previous discourse, common ground, and external surroundings:

The sentence ‘\( a \) believes \( b \) to be \( \varphi \)’ uttered in context \( C \) is true iff there is a description \( \alpha \) suitable in \( C \) such that \( \alpha \) is actually \( b \) and \( a \) believes that \( \alpha \) is \( \varphi \).

\([\text{Aloni} 2000:61]\)

Aloni discusses a number of logicians and philosophers (Kraut 1983; Gerbrandy 2000) who similarly emphasize the role of the context in analyzing de re belief under a description, eventually tracing the idea of quantification over contextually selected concepts back to Hintikka (1967; 1969). Seemingly independently, linguists working in the Lewis-Cresswell-Von Stechow tradition have come up with the same ideas within the relational attitude framework. In the current section we explore a couple of ways to formalize, and benefit from, this pragmatic aspect of de re (and de se) belief.

---

12A closer look at our reconstruction of the vivid name account reveals that the context dependence of res descriptions was already, though less explicitly, captured in the vividness and acquaintance criteria within our earlier definition of \( R \), (30), p. 33.

13(Aloni 2000:fn.2 p.35 & fn.34 p.57)
2.3.2.1 Contextualized acquaintance

Let’s heap together the concept quantification, vivid names and relational analyses, and discuss some of the ways in which they have been contextualized and why. First up is an extension of the relational framework by Abusch (1997) who argues that the acquaintance variable \( R \) should be determined by the context rather than existentially quantified over as in (11). Her motivation is twofold: (i) a contextually determined \( R \) can be kept constant through a sequence of belief reports. For instance, if in the context of a particular discourse, Ortcutt is presented to Colby as the philosopher with the blue hat over there, then every \( de \ re \) ascription about Ortcutt to Colby will pick up that description. This comes closer to what actually happens when we interpret such a discourse than looking for the existence of some acquaintance relation in each sentence anew. However, I doubt that this will make a truth conditional difference, because the quantification over acquaintance or descriptions in the original proposals (Kaplan 1969; Cresswell & von Stechow 1982) is restricted to suitable, vivid ones, and as such already anticipates that the relation must be salient in the context.

Second, Abusch claims her analysis allows the possibility of judging a report true in one context, yet false in another. For instance, if I had cut short the Ortcutt story halfway, only telling the audience that Janell admires Professor Ortcutt, then the report “Janell believes Ortcutt is a genius” could be true, whereas “Janell believes Ortcutt is a loser” would be false. If I were to focus entirely on the second half, about how Janell hates the abstract that is actually written by a professor Ortcutt, these judgments would be reversed. Related to Abusch’ second point I see another potentially big advantage of contextualizing acquaintance in the fact that the whole \( de \ re/de \ se \) distinction is now a matter of pragmatic ambiguity. A given ‘coreferential’ report, i.e. of the form \( NP_a \) believes \( NP_a \) VP, is pure \( de \ se \) iff the context’s most salient acquaintance relation is equality \( (a = a) \), otherwise it can be intentional \( de \ se \), mere \( de \ re \), or infelicitous, depending on what relations the context has to offer (from those enumerated in table 1.1, p. 54).

In short, I agree wholeheartedly with Abusch’ motivations for going contextual. In fact, I think the contextual/pragmatic aspect of \( de \ re \) belief should be stressed even more, as will become apparent in the remainder of this chapter and the next. Playing devil’s advocate though, I might again just point to Kaplan’s suitability/vividness clause and say that this contextuality aspect of \( de \ re \) was already well taken care of in the sixties. Now, rather than these implicit allusions to contextuality inherent in vague notions like vividness, what we want is a theory about the nature of this context dependence, and how to represent and resolve it. Unfortunately Abusch offers
nothing in the way of a formal account of an underlying pragmatics, she merely leaves room for one by dispensing with the whole $\exists R[R(a, b) \wedge ]$ that we saw in relational lfs. Her contextualized relational analysis thus captures the semantics of de re as follows (Abusch 1997:ex.13):

(20) $\text{NP}_a$ believes of $\text{NP}_b$ that it is $\text{VP}_P$

$\triangleright \text{BEL}_a^x \lambda x [P(\forall y[R(x, y)])]$

[cf. (11), p. 80]

The description under which the belief is held is represented by a free relation variable, $R$, which an independent and as yet unspecified pragmatic module is supposed to bind to a contextually salient relation. An important restriction on the process of identifying a suitable $R$ is obviously the fact that $R$ should hold between the ascribee and the res, which was hardwired into the original account but is invisible in the new lf, presumably relegated to the aforementioned pragmatic module. Pending a formalization of such a module this could be a pro or a con. What the Abusch treatment does not accomplish however, is making the definition of de re more compositional: it merely throws the arguably problematic wide scope acquaintance quantifier into the 'pragmatic wastebasket', which never solves anything. More on de re compositionality issues in 2.3.3.

Let’s zoom out and consider how this free variable trick, which is ubiquitous in a certain cluster of syntax-semantic interface theories, actually works. To interpret a formula like (20), we have two options: besides seeing $R$ as a free variable that has yet to be bound by a quantifier ‘in the discourse context’, we can interpret $R$ as an indexical constant, receiving an interpretation from the context parameter of a two-dimensional semantics with separate context and index parameters. The first option seems ill-fitting in a static framework like Abusch’s because it presupposes a pragmasemantic theory where contextual information is stored in formulas that are expanded when new sentences are interpreted, thus allowing existential quantifiers in the context representation to bind free variables in subsequent utterance representations.

14 (Bar-Hillel 1971)

15 A third option, mixing the open formula account with the indexicality aspect, is to stipulate that free variables should get an interpretation directly from the assignment function, which in turn is determined by the context. This is how for instance Schlenker (2003) handles demonstratives. To make it work, a new aspect of interpretation is introduced, the “compatibility condition”, which states that a logical form should be interpreted w.r.t. a proper assignment determined by the context, i.e. the assignment should map the formula’s free variables to appropriate individuals from the context, in accordance with the speaker’s referential intentions in that context. The insider may note the strong similarity with the standard DRT treatment of rigidity, in terms of external anchors (to be discussed in 3.3.2.3).
All this has a much more natural formulation in dynamic semantics, perhaps with the aid of a presuppositional mechanism for binding the free R to an appropriate contextual relation ... (yes, this is my analysis in a nutshell, details in chapter 3)

The second option, where R is an indexical constant, seems to be in accordance with Aloni (2000), who clarifies the pragmatic account of de re belief thus:

The idea behind the pragmatic analysis is that de re belief reports express different contents in different contexts, in the same way (or in a similar way) as sentences containing indexical expressions.  

[Aloni 2000:61]

Apart from the fact that Aloni wants to contextualize descriptions/guises instead of 2-place relations of acquaintance, she seems to be aiming at the same kind of result as Abusch. The main difference would be that Aloni’s view requires a two-dimensional semantics that allows one to evaluate R on a context parameter (provided by the model), rather than a (semi-)dynamic semantics where context representations are ‘updated’ with incoming sentences.16

Besides the advantages of de re contextualism already mentioned, both Abusch and Aloni-style implementations of it share one important and distinguishing prediction: the vivid acquaintance relations holding between matrix and embedded subjects are interpreted wide scope with respect to further operators embedding the belief report. Consider a quantified belief report like Every student believes she’ll win. Cresswell & von Stechow’s (1982) definition, our (11), would assign it a reading where the acquaintance quantification is built into the semantics of believe, so that it is outscoped by the universal quantifier. The resulting reading is paraphrasable as for each student there exists an acquaintance relation R between that student and herself, and that student believes under that R that. . . . The context dependent extensions described above (whether employing pragmatic discourse binding, free variables, or indexical R’s)17 predict a truth-conditionally distinct read-

16In fact, Aloni’s technique is still slightly different: instead of making the description itself an indexical, she has a contextually determined set of concepts, which the If then quantifies over. One difference with analyzing R as an indexical constant is that it then carries a kind of built-in uniqueness presupposition, whereas Aloni’s contexts need only determine a set of concepts. It remains to be seen if anything hinges on this superficial difference. Note that with an Aloni-style definition, Abusch would no longer be able to sweep the above mentioned non-compositionality under the carpet, since the existential quantifier must remain visible at If.

17I’m not sure how well Aloni’s (2000) version (footnote 16) works. Perhaps the flexibility in choosing the contextual set will allow her to derive both wide and narrow scope readings in principle.
2.3 De se and de re

There is a contextually determined acquaintance relation $R$ s.t. each of the students bears it to herself and believes under $R$ that... We return to this matter in 2.4.2, where we take a closer look at belief reports embedded under quantifiers (esp. every and only). I show there that neither existential closure nor simple (rigidly wide scope) context dependence will do, thus clearing the way for a more sophisticated, truly pragmasemantic treatment of acquaintance relations.

2.3.3 Compositionality?

Some recent authors have expressed their dislike of the relational analysis and its Kaplanian predecessor on the grounds of its apparently obvious non-compositionality. For instance, Schlenker (2003) talks about “the notorious difficulty of giving a compositional treatment of Kaplan’s suggestions” (p.94), while von Stechow & Zimmermann (2004) describe that account as “Kaplan (1969)’s classical but non-compositional de re account” (p.15). Although both remarks are overtly aimed at Kaplan, not at the relational generalization we’ve ascribed to Cresswell & von Stechow (1982), these qualms immediately extend to that account (and its contextualizations).

The source of the non-compositionality lies in the fact that $res$ and ascribed predicate must be kept apart throughout the semantic derivation. For instance, say we want to interpret Colby believes that Ligia is smart compositionally, i.e. computing the meaning of the whole sentence as a function of the meanings of the constituents, and the meanings of those constituents in terms of their constituents, and so on, down to the lexical items (cf. 1.1.1.3).

This means that we must be able to compute the semantic value of the complement clause (a constituent on all accounts) first, as if it were a sentence on its own. This would give us the proposition that Ligia is smart. Note that from this proposition, a plain set of worlds, we can no longer recover the individual Ligia or the property of being smart. But then how do we continue our computation? Well, in the next steps the values of Colby (an individual) and believes (an operator) must combine with this proposition, which leads to a de dicto interpretation at best. The de re truth conditions, $\exists R[c, 1] \land Bel^c_x[smart(ry[R(x, y)])]$, are unattainable since they feature both Colby and Ligia in the main, unembedded context, implying that Ligia

---

18. This is the most charitable reading I can imagine on a wide scope analysis. I’m not sure how Abusch in particular could derive the requirement that $R$ holds between the student and herself in narrow scope, while $R$ itself takes wide scope.

19. Common enhancements to this principle of compositionality include adherence to a particular independent syntactically motivated tree structure that the semantics must follow, or a restriction to functional application as the only mode of semantic composition.
was available at some point after the computation of the complement had been completed. As if by “magic” (Schlenker 2004b), the res must have ‘moved’, from the embedded to the main clause, a phenomenon referred to as res movement by von Stechow & Zimmermann (2004). In any case, the above should suffice as a proof that we cannot derive relational lfs by purely compositional means.

In order to evaluate these recent attacks based on non-compositionality, let us take a look at how that same Arnim von Stechow, in collaboration with Max Cresswell, had worked out the relation between surface structure and the relational logical form in the eighties, for it is at that interface that compositionality appears to be problematic. The alternatives proposed by von Stechow & Zimmermann (2004) and Schlenker (2003; 2004b) on the basis of their criticism of the relational analysis will be taken up later, in 2.3.4 and 2.4.3.

2.3.3.1 Structured propositions in logical forms

Cresswell & von Stechow (1982) use a framework in which the LF, generated by a categorial syntax à la (Cresswell 1973), of a de re belief complement is a structured proposition. For example:

(21)  Ellsworth believes of Arata Suggs that she’ll win \([=(1b)]\)
\[\triangleright \text{BEL}_x(a,\text{win})\]

This is in accordance with the remarks about structured proposition we made when talking about the semantics of de re believing in 1.2.2.1. There we observed that de re belief calls for singular propositions, and the most straightforward way to formalize the aboutness of singular propositions was in terms of structured propositions that keep res and predicate separate. Hence this idea of modeling de re in terms of structured propositions, especially when the aboutness is as explicit as in the \(x \text{ believes } P \text{ of } y\) construction.

The same analysis can be applied to the de re reading of our ambiguous report, which is now ambiguous in the sense of admitting the lfs in (22a,b).

(22)  Ellsworth believes the prettiest girl will win \([=(2)]\)
   a. \(\text{BEL}_x\text{win}(\forall x.\text{prettiest}(x))\) \([=(13a)]\)
   b. \(\text{BEL}_x\langle x.\text{prettiest}(x),\text{win}\rangle\)

Cresswell & von Stechow (1982) in fact derive this ambiguity from a lexical ambiguity in the (sometimes invisible) complementizer that, which on their account can select for complements of various logical types, which include propositions and (subject,predicate) structured propositions as in
(22a) and (22b), respectively. The choice of a particular member of the *that* family determines whether, and in what way, the structure of the complement is relevant for interpretation, in other words, which terms are to be read *de re*: in (22b), only the subject term is *de re*, in (22a) nothing is. In the following I will not follow Cresswell’s *that*-centrism, but locate the ambiguity in the BEL operator.

In general, a belief complement’s If has the form \( \langle \langle \alpha_1, \ldots, \alpha_n \rangle, \beta \rangle \), where the \( \alpha \)'s are the terms (of arbitrary type, so we will allow reports that are *de re* about, say, properties) read *de re*, and \( \beta \) is the property believed to hold of these objects. The logical types of the \( \alpha \)'s and \( \beta \) are restricted by the requirement that applying the predicate \( \beta \) to the \( \alpha \)'s yields a (normal) proposition:

\[
\langle \langle \alpha_1, \ldots, \alpha_n \rangle, \beta \rangle \text{ is a structured proposition iff } \beta(\alpha_1, \ldots, \alpha_n) \text{ is a well-formed formula}
\]

Given the notational convention that we may leave out superfluous angled brackets, i.e. \( \langle \alpha \rangle = \alpha \), it’s easily seen that this holds in example (22), where we had the limiting cases of \( n = 0 \) (*de dicto*) and \( n = 1 \) (\( \langle \langle x[\text{prettiest}(x)] \rangle, \text{win} \rangle \) is a structured proposition because \( \text{win}(x[\text{prettiest}(x)]) \) expresses a regular proposition).

As I showed at the start of 2.3.3, a separation of *res* and predicate is necessary even though it has the effect of making things non-compositional. We now see how structured propositions come in handy for implementing this *res* separation. The term *res movement* is especially apt when we turn to slightly more complex cases like (24). Consider for instance the reading where it’s the transitive verb’s object that is read *de re*, in fact, to demonstrate the power of the system, let me show you three such readings, for in addition to the object we can ‘*res*-move’ the other embedded constituents as well:

\[
(24) \quad \text{Ellsworth thinks the prettiest girl might beat Arata}
\]

\[
a. \quad \text{BEL}_e\langle \langle a \rangle, \lambda y[\circ \text{beat}(x[\text{prettiest}(x)], y)] \rangle
\]

\[
b. \quad \text{BEL}_e\langle \langle a, x[\text{prettiest}(x)] \rangle, \lambda y \lambda z[\circ \text{beat}(z, y)] \rangle
\]

\[
c. \quad \text{BEL}_e\langle \langle a, x[\text{prettiest}(x)], \text{beat} \rangle, \lambda y \lambda z \lambda x[\circ \text{X}(z, y)] \rangle
\]

Already in the simplest If, (24a), in which the belief is *de re* only with respect to the individual called Arata, it’s clear that the *res* has been “move[d] to the left edge” of (the LF of) the embedded clause “thereby creating a \( \lambda \)-abstract” (von Stechow & Zimmermann 2004:3). And if it wasn’t already, it should now become obvious that there’s no way to get from *the prettiest girl will beat Arata* to \( \langle \langle a \rangle, \lambda y[\circ \text{beat}(x[\text{prettiest}(x)], y)] \rangle \) in a compositional, non
This then is the compositionality problem for the relational account. A counterargument might take the following form: If movement makes a syntactic/semantic theory uncompositional, then apparently your notion of compositionality is way too strict, for there is ample evidence in syntactic theory that movement is indispensable. The hard-core Montagovian standpoint regarding compositionality is really untenable in light of the actual syntactic complexity of natural language. Besides, movements occur in syntax, not semantics. It comes as no surprise that compositionality doesn’t apply to the relation between the surface form and the interpretation, what matters is that it applies between the LF (after the movements) and its model theoretic interpretation. This should be an acceptable solution, since indeed it is (or certainly was) uncontroversial among the majority of syntacticians that movements are needed. The remaining questions are then whether this particular movement is well-motivated on independent syntactic grounds, and whether a structured LF can indeed be interpreted compositionally according to the relational attitude analysis. As for the first question I remain agnostic, deferring to von Stechow & Zimmermann (2004) who note there has been some controversy about res movement, but, in their eyes, it’s safe, since the syntax/semantics of focus requires the same “long” movement.

To answer the second question, whether we can compositionally interpret structured LFs, we have to look a bit deeper into the theory. Consider again the examples in (24). Strictly speaking, since the lf complements of those beliefs are all of different types, the operators must be different as well, which would be in accordance with the idea that the complementizer, or, as we prefer it, the belief operator itself, is multiply ambiguous. We should really distinguish all these Bel’s, e.g. by superscripting the types of its complements: Bel_{et} et in (24a) and Bel_{et} et in (24b). We have not really introduced syntactic LFs, but we assume that their derivation involves res movements so that they are isomorphic to these structured formulas. To show how such LFs are compositionally interpretable as having relational truth conditions, let me simply define the new structure sensitive Bel items in such a way that formulas containing them are just abbreviations of relational lfs:

\[ \text{If } \tau_0 \text{ and } \tau_1 \text{ are types s.t. } \tau_0 \times \tau_1 \text{ is the type of a structured proposition (cf. (23) and 2.B), then:} \]

\[ (25) \]
2.3 De se and de re

\[ \text{BEL}^{\text{a}}_{\text{et}} \equiv \lambda p,\text{et} \exists R_{\text{et}}[R(a, (p)\text{et}) \wedge \text{BEL}^{\ast}_{\text{et}} \lambda x[((p)\text{et})(\text{et}x, [R(x, y)])] \]

More concretely, take (24a):

\[
\text{(26) Ellsworth thinks the prettiest girl might beat Arata}
\]

\[ \text{BEL}^{\text{et}}_{\text{et}}(a, \lambda y[\text{beat}(\text{et}x[\text{prettiest}(x)], y)]) \equiv (\lambda p,\text{et} \exists R_{\text{et}}[R(e, (p)\text{et}) \wedge \text{BEL}^{\ast}_{\text{et}} \lambda z[(((p)\text{et})(\text{et}z, [R(z, w)])] \]

\[ \exists R_{\text{et}}[R(e, a) \wedge \text{BEL}^{\ast}_{\text{et}} \lambda z[(((p)\text{et})(\text{et}z, [\text{beat}(\text{et}x[\text{prettiest}(x)], y)])] \]

Simply applying the definition of \text{BEL}^{\text{et}}_{\text{et}} from (25) we thus derive the result that structured belief LFs (or LFs) can be seen as abbreviations of relational truth conditions. This concludes the proof that the non-compositionality is real, but can be relegated to syntax (\text{res} movements at LF): once we have structured complements separating \text{res} and ascribed predicates we can proceed compositionally, with just functional application. A disadvantage of the approach may be the necessity of an infinitely ambiguous belief operator, but accepting that, everything goes rather smoothly.

Of course, a new dimension of non-compositionality will be introduced if we allow pragmatic binding of free variables in the context, as proposed in 2.3.2. This really can’t be helped, but, as we have attempted to push \text{res} movement into syntax, so we might also say that this new form of non-compositionality belongs to pragmatics. Compositional semantics is saved by two wastebaskets. In the next section we’ll first explore whether we can get around the syntax/semantics compositionality issues. The pragmatic aspects will investigated later.

2.3.4 Kaplan’s compositional alternative

Discussing the relational analysis just described, von Stechow & Zimmermann (2004) conclude that, although such an account can give accurate enough predictions, it is semantically and syntactically “much more complicated” (p.3), requiring ad hoc transformations, which are at the root of the “non-compositional” (p.15) nature of the analysis, further remarking that “[\text{linguists could object that ‘res movement’}, i.e., the movement of \text{[the embedded subject] from the specifier position to the left boundary of the sentence violates well-established locality restrictions and is not attested elsewhere in syntax.” (p.3). For these reasons, they consider it time to in-
vestigate the merits of an alternative account, one that has been circulating in philosophical circles for decades, but has not been very popular in linguistics, which is remarkable perhaps, for it certainly seems to beat the relational analysis on compositionality by interpreting all res in situ. This section introduces the compositional alternative, Kaplan’s (1989) so-called ‘Adding “Says”’ analysis, as a further development of the Kaplanian two-dimensional semantics.

In the final subsections of chapter 1 we compared the relational analysis of belief to Kaplan’s (1989) two-dimensional analysis, according to which the object of belief is (the diagonal of) a character. In 1.2.4.2 and 1.B we followed Kaplan (1989:ch.XVII) in explicating belief as a relation between individuals and characters, where characters were two-dimensional generalizations of the notion of a proposition: a sentence’s character maps every context of utterance to the proposition expressed in that context. Formally, contexts function as a second semantic parameter, so that every expression gets a denotation only with respect to a certain context and a possible world.

The idea behind Kaplan’s attitude semantics was that a character encodes all the semantically relevant information carried by a sentence, and is thus finegrained enough to analyze purely indexical distinctions. And that is what’s needed for an analysis of mental content, because even if he’s skinny and I’m skinny express the same proposition (in other words, if the distinction is merely indexical, not truth-conditional, as when the first is uttered in a mistaken self-identity scenario), the underlying (narrow) attitudes are quite far apart. As for belief, it was shown in 1.2.4.3 that we can improve upon Kaplan’s characterial belief, following common post-Kaplanian practice, by restricting attention to diagonals. From appendix 1.B we now repeat the diagonal version of Kaplan’s semantics of belief reports:

\[
\text{(27) Belief (diagonal version of Kaplan (1989:ch.XVII))} \quad [\approx (72), \text{p. 69}]
\]

a. Syntax: If \( \tau \) is a term and \( \varphi \) a formula, \( \text{bel}_\tau \varphi \) is a formula

b. Model: \( \text{Bel}^\mathcal{C} : D \times \mathcal{I} \rightarrow \wp \mathcal{C} \)

c. Semantics:

(i) \( \llbracket \varphi \rrbracket^f = \delta \llbracket \varphi \rrbracket^f = \left\{ c \in \mathcal{C} \mid \llbracket \varphi \rrbracket^f_{(w_i,t_i)} = 1 \right\} \) \[\approx (70b), \text{p. 69}\]

(ii) \( \llbracket \text{bel}_\tau \varphi \rrbracket^f_{c} = 1 \text{ iff } \llbracket \varphi \rrbracket^f \supseteq \text{Bel}^\mathcal{C}(\llbracket \tau \rrbracket^f_{c}, i) \)

What would happen if we took this simple belief semantics as a semantics for belief reports? Recall that for the relational framework such a move worked reasonably well, the semantics of belief could be used immediately as a semantics of belief reports. In 1.2.4.4 we examined how that framework and Kaplan’s relate to each other with regard to belief and we concluded that they had about the same empirical adequacy, so let’s follow the relational
analysis’ lead and try to interpret (27) as a semantics of reports.

To see what happens, consider a report with an embedded indexical, like *me* in (28) (represented as *i* in the if):

(28) Arata thinks the dumbest girl can beat me

\[ \text{\text{BEL}_a \text{can_beat}(\text{\text{x.dumbest}(x)}, i)} \]

Note the nicely compositional (in situ) interpretation of *me*, which contrasts with the relational account that has to move it to the front of the embedded clause first, as in (24). Following (27), we interpret the if in (28) as in (30), for which we first need to compute a diagonal, (29):

(29) \[ \llbracket \text{can_beat}(\text{x.dumbest}(x), i) \rrbracket^f \]

\[ = \{ c \in C \mid [\text{can_beat}(\text{x.dumbest}(x), i)]^{f,c}_{(w,c)} = 1 \} \]

\[ = \{ c \in C \mid \langle \llbracket \text{x.dumbest}(x) \rrbracket^{f,c}_{(w,c)}, [i]^{f,c}_{(w,c)} \rangle \in \llbracket \text{can_beat} \rrbracket^{f,c}_{(w,c)} = 1 \} \]

\[ = \text{the set of contexts whose agent can be beaten by the dumbest girl of that context’s world and time} \]

(30) \[ [\llbracket (28) \rrbracket]^{f,c}_i = 1 \]

iff \[ \llbracket \text{can_beat}(\text{x.dumbest}(x), i) \rrbracket^f \subseteq \text{BEL}^{f,c}(\llbracket a \rrbracket^f_i, i) \]

iff in all of Arata’s belief alternatives the center can be beaten by the dumbest girl

iff Arata believes: “I can be beaten by the dumbest girl”

This is definitely not a reading the sentence actually has, because in real life *me* refers to the actual utterer, not to Arata (or the person she believes to be, i.e. an agent of one of her belief alternatives). Moreover, this systematically incorrect prediction with respect to indexicals is not a result of the post-Kaplanian shift from characters to diagonals, as the hardcore Kaplanian skeptic could verify, e.g. by replacing \text{BEL} in (29)-(30) with \text{SAY}, the strictly Kaplanian operator defined in (75bi), p. 70. It’s really the in situ interpretation of belief-embedded terms that is wreaking havoc.

But Kaplan has a solution, based on the following reasoning: Indexicals are rigid designators (“Principle 2” (Kaplan 1989:492), cf. 1.2.3.3), so their reference cannot be shifted by natural language operators (“Fixity Thesis” (Schlenker 2003:29), cf. 1.2.4.2), in other words, natural language operators are at most intensional. This last constraint is often called “the prohibition of monsters”, because Kaplan (1989:510) coined the term “monsters” for characterial operators, i.e. operators, all too easily definable in his formal system, that need more than their argument’s actual intensions to give an output. In fact, the operator \text{BEL} above is a good example of a monster,
since it depends on the diagonal expressed by its argument, which is more, or rather to some extent independent of the intension actually expressed. Kaplan’s original indirect speech semantics, (72), p. 69, was even worse, taking into account the argument’s entire character, modeling the reported speech’s or attitude’s linguistic and/or cognitive significance.

It’s easily verified that a monster like \texttt{bel} indeed violates Principle 2; we just saw how the diagonal shifts a \texttt{bel} embedded indexical’s reference in the computation in (29). This is not yet a real problem, because Kaplan only forbids monsters in natural language, not in the analysis of attitudes. In fact, we were already looking for a different analysis of the linguistic operator \texttt{believes (that)} because we just showed that it gives the wrong predictions, at least for English reports with embedded indexicals. We have to come up with a semantics where interpreting a complex formula at a context and an index depends only on the intensions of the parts with respect to that same context. To decide if (28) (Arata thinks that the dumbest can beat me) is true in \(c\), we should not have to look any deeper than at the proposition actually expressed by the complement clause in \(c\), which in this case is that the dumbest girl can beat Emar. The characterial or diagonal sensitivity in the definition of belief, (27), was meant to give the exact mode of presentation under which a belief is held, so in an intensional system this will have to be abandoned, e.g. as Kaplan suggests for indirect speech, by “quantifying away” the characterial mode of presentation of the direct speech. In his own words, \(x\) says that . . . is true if:

\[
\exists c, C \; [c \text{ is a context } \land \; C \text{ is a character } \land \; x \text{ is the agent of } c \land \\
\text{x direct-discourse-verb } C \text{ at time } t \text{ of } c \land \text{ the content of } C \text{ in } c \text{ is that . . . } ]
\]

[(Kaplan 1989:554)]

In other words, for a report to be true there has to be some character \(C\) (of the original utterance) and context \(c\) (the original speech context) such that \(C \in Say \text{ in } c\), and the original’s propositional content in its utterance context \((c) = \text{ the report complement’s propositional content in its utterance context. The only level of content needed from the “...”—the complement of the indirect speech operator—is the proposition actually expressed by it. The rest of the character of the complement is irrelevant; the requirement that there is some original character (usually different from the complement’s character because of speaker or time changes) that was directly said doesn’t change that. For this reason, the prohibition on monsters is not violated, and consequently Principle 2 is unharmed. In short, since the complement is only evaluated in the actual report’s utterance context, any indexical in it will receive its unshifted reference.
2.3 De se and de re

In appendix 1.B.4 we formalize and extend this analysis of indirect discourse to belief reports. The main difference with the Kaplanian original is that for belief, we argued in 1.2.4.3, full character would be overkill, we are much better off using just diagonals, i.e. (27). In other words, we define believing a character as self-ascribing the diagonal, which is further reducible to the familiar $\text{Bel} \subseteq \delta C$. One consequence for the resulting report semantics, constituting a major departure from Kaplan’s original sketch of indirect discourse, is the fact that we now have a kind of inferential closure for belief and thus for reports. In any case:

\begin{align}
(31) \quad [\text{Bel}^\text{mar}_r \varphi]^f_c &= 1 \text{ iff there is a character } C \text{ s.t. } \delta C \supseteq \text{Bel}^C([r]^f_c, i)
\text{ and } [\varphi]^f_c = C([r]^f_c, t_i, w_i) \tag{\approx(75), p. 70}
\end{align}

Returning to our earlier example (28), it will become evident how (31) achieves its goal: a semantics that can handle de re and de se belief reports uniformly and in a non-monstrous way. Moreover we can keep the same clean if with the in situ indexicals, replacing only $\text{Bel}$ with $\text{Bel}^\text{mar}$. Crucially, computing the truth conditions according to (31) no longer requires the computation of a diagonal or other monstrosities. To get at the truth conditions, let’s assume that I now say (28) to you, i.e. $a_c = \text{Emar}$, then the Kaplanian analysis predicts:

\begin{align}
(32) \quad [\text{Bel}^\text{mar}_r \text{can}_c \text{beat}_c r \text{. dumbest}_c (x), i]^f_c &= 1 \\
&\text{ iff there is a } C \text{ s.t. } \delta C \supseteq \text{Bel}^C([a]^f_c, i) \text{ and } \[\text{can}_c \text{beat}_c r \text{. dumbest}_c (x), i]^f_c = C([a]^f_c, t_i, w_i) \\
&\text{ iff there is a } C \text{ s.t. } \delta C \supseteq \text{Bel}^C(\text{arata}, i) \text{ and } \{j \mid [r\text{. dumbest}_c (x)]^c_j, [i]^c_j \in \mathbb{I}(\text{can}_c \text{beat}_c)(c)(j)\}
= C(\langle \text{arata}, t_i, w_i \rangle) \\
&\text{ iff there is a } C \text{ s.t. Arata self-ascribes } \delta C \text{ in } i, \text{ and } C(\langle \text{arata}, t_i, w_i \rangle) = \text{ the proposition that the dumbest girl can beat Emar.}
\end{align}

In the above example, the belief complement actually expresses the proposition that Emar can be beaten by the dumbest girl, the report context is the current index $i$ with Arata as center, so $C$, the original thought can be for instance the character of any sentence of the form the dumbest girl can beat $x$, where $x$ rigidly refers to Emar in Arata’s context. Even more concrete, (28) counts as a report of Arata’s sincere utterance of any of (33), even though these sentences have very different characters and correspond to (are expressions of) indexically distinct attitudes of Arata’s:
The dumbest girl can beat \{ you [addressing Emar] \\
that guy [pointing at Emar] \\
him [anaphoric to Emar] \}

Let us look at some more concrete predictions to see how far this theory goes. In fact, let’s tackle our familiar test cases from 2.1. For the de dicto report (1a) we get the following truth conditions:

$$\text{(34) Ellsworth Kimmel believes pretty girls always win} \quad [\approx (1a), \text{p. 72}]$$

$$\Rightarrow \text{BEL}_{e^{\text{true}}} \forall x[\text{pretty}(x) \rightarrow \text{win}(x)] \quad [\approx (4)]$$

$$[[\forall x[\text{pretty}(x) \rightarrow \text{win}(x)]]_{f,c}^{e} = 1 \text{ iff } [\text{by (31)}] \text{ there’s a } C \text{ s.t. } \delta C \supseteq \text{BEL}(\text{ellsworth}, i) \text{ and }$$

$$[[\forall x[\text{pretty}(x) \rightarrow \text{win}(x)]]]_{f,c}^{e} = C((\text{ellsworth}, t_i, w_i))$$

These truth conditions amount to saying that there is a character/utterance/thought expressing a belief of Arata’s which has the same content as the report complement, i.e. that all pretty girls win. Given that there are no indexicals or other context dependent terms in there, the original utterance/thought may well have been totally the same, characterially, (“Pretty girls always win” or a translation thereof) though it need not be, since the original’s full character is heavily underdetermined (a very important point that shall be taken up later). One a priori problem of this underdetermination is that it is quite hard to prove that a report really is false, because that’s a negative existential statement: there is no character that fulfills... In computing truth values, for now, we’ll focus on saying when and why some reports are obviously true in some familiar contexts, by providing a verifying character.

We want to compare the current theory’s predictions with those of the rival (partial) accounts of the data represented in table 2.1, that is to say, the scopal accounts of de dicto/de re in 2.2.1, the Kaplan (1969) account of de re in 2.2.2, and the relational account of de re/de se reporting from 2.3.1. Thus, the contexts we are interested in are the by now familiar ones from 2.1, A, B, and C, so in order to get concrete truth values to fill the schema, let \(c_a\), the one uttering the report, be me, and let the time and world of context \(c\) and index \(i\) be determined by the situations described in A through C. For our first If (34) that gives truth at A with verifying character \(C_1 = [\forall x[\text{pretty}(x) \rightarrow \text{win}(x)]]\). We want to get falsity for B and C, and indeed \(C_1\)’s diagonal is not self-ascribed in these contexts, so at least that character won’t verify it. Whether there really is no verifying character remains to be seen.

Next, the de re report:
2.3 De se and de re

(35) Ellsworth believes of Arata that she’ll win

\[ (35) \] \[ f,c \] = 1 iff there’s a C s.t. \( \delta C \supseteq B_e C(ellsworth, i) \) and

\[ \text{[win(a)]} f,c = C(\langle ellsworth, t, w \rangle) \]

Here, a is a rigid designator, but still not context dependent (let’s assume that much). As a result, (35)’s truth conditions are fulfilled by \( C_2 = \text{[win(a)]} \), i.e. in B, and by \( C_3 = \text{[win(i)]} \), in C. We leave A, where we hope to get falsity, for later.

Thirdly, the de dicto/de re ambiguous report. First consider what happens on the most obvious If, the one with an in situ definite description:

(36) Ellsworth believes the prettiest girl will win

\[ (36) \] \[ f,c \] = 1 iff there’s a C s.t. \( \delta C \supseteq B_e C(ellsworth, i) \) and

\[ \text{[win(a)]} f,c = C(\langle ellsworth, t, w \rangle) \]

So construed, the report should be de dicto, and indeed, taking \( \text{[win(x.prettiest(x))] for } C_4 \) will do the trick. For a de re if we have two options: scope out the description as in 2.2.1, or somehow rigidify it. The result will be the same, but since Kaplan’s system introduces a special operator, DTHAT, for rigidification, let’s see how that works. In 1.B we defined DTHAT as follows:

(37) [DTHAT e] f,c = \( e \) f,c \( \langle w_c, t_c \rangle \)

Applied to the embedded description that makes:

(38) \( B_e \text{[win(DTHAT(x.prettiest(x)))]} \)

And then:

\[ ([38]) f,c = 1 \text{ iff there’s a } C \text{ s.t. } \delta C \supseteq B_e C(ellsworth, i) \text{ and } \]

\[ \{ j \in I \mid \text{[x.prettiest(x)]} f,c (w_c, t_c) \in \text{[win]} f,c \} = C(\langle ellsworth, t, w \rangle) \]

In other words, the prettiest is now evaluated only in the actual context, not in subjective belief alternative indices, and therefore this If is equivalent to a wide scope one, and even, since in all of A through C Arata = Ellsworth = the prettiest, to the one in (35).

Postponing the unambiguously pure de se report, let’s first deal with the embedded co-referential pronoun, since that raises one of the issues Kaplan himself spends some time on, and which he considers one of the main motivations for his report semantics:
Ellsworth believes she’s going to win

\[\text{BEL}^\text{true}_{e}\text{win}(e)\]

This If is satisfied iff Ellsworth believes some character that expresses the proposition that Ellsworth, under whatever mode of presentation, will win. In other words, the report is true if the reported belief were pure, 1st person *de se*, i.e. \(C_3\) ("I am going to win"), but also if the belief were expressed under a mere *de re* or intentional *de se* guise like \(C_2\) ("Arata will win") or "that woman [pointing at self on TV] will win" respectively. This means that, just as on the relational account no ambiguity is needed to account for the judgment that (39) is true in both pure (C) and impure (B) *de se* scenarios. In this sense, both analyses offer a unified account of *de se* reports, while still powerful enough to represent the distinct truth conditions of the various types of real *de re/de se* beliefs. In the relational account we distinguished the subtypes by classifying acquaintance relations (table 1.1, p.54), an analogous classification here would require a typology of characters or diagonals believed.

This smooth account of (39) however has an immediate drawback analogous to the one we encountered with the earlier unificatory proposal of relational attitudes. Both unified accounts of *de re/de se* reports predict that language does not distinguish between pure *de se* and other *de re* beliefs. In the current system this is because a single If is compatible with any mode of presentation as long as the expressed proposition matches with the report’s complement’s. This should come as no surprise for it’s very explicitly built into Kaplan’s system by means of the prohibition on monsters, which, formulated as a composition principle, says that the semantic value of a belief report depends only on the intensions of its component parts (subject NP and complement S). The problem then is, how can there be reports like (1c) that are true only in pure *de se* scenarios? Let’s see what we really predict for (1c). First of all, we mold it in the prescribed syntactic format by assuming an (independently motivated) PRO as subject, which we then assume to refer to the matrix subject, rigidly, just as in 2.3.1.

Ellsworth believes PRO to be on the winning side

\[\text{BEL}^\text{true}_{e}\text{win}(e)\]

I claim that this is the best we are going to get on a Kaplanian analysis. Unfortunately it turns out to be the same If as in (39), which won’t do, because that one is compatible with any *de re* scenario, in particular one where Ellsworth ascribes the winning to her mirror image that she doesn’t recognize.
As before we now gather the predictions discussed above and display them in a table, 2.5, for comparison with our intuitions in table 2.1. I have put question marks at all coordinates where we have not found verifying characters representing modes of presentation occurring in the story contexts, but we have not yet shown that there really aren’t any. The results so far

<table>
<thead>
<tr>
<th></th>
<th>A.</th>
<th>B.</th>
<th>C.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>(34)</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>1b</td>
<td>(35)</td>
<td>?</td>
<td>1</td>
</tr>
<tr>
<td>1c</td>
<td>(40)</td>
<td>?</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.5: Preliminary predictions of Kaplan’s (1989) ‘Adding “Says”’

(with tentative zeros at the question marked positions) coincide on the *de dicto/de re* part with the analyses of 2.2, and on the *de re/de se* part with the relational analysis of 2.3.1. The main problem already rearing its head in this preliminary table is the pure *de se* report (1c), which however was equally problematic for the relational analysis, so we’ll discuss that particular problem and some solutions in 2.4.

### 2.3.5 Kaplan refuted

Both reductionist attempts, relational attitudes (with acquaintance relations) and Kaplan’s (1989) no-monster theory above, fail with respect to pure *de se* reports, and more or less for the same reasons too. As for the other belief reports discussed so far, the preliminary predictions coincide as well. Theoretically, though, the Kaplanian character theory has the big advantage of being fully compositional, whereas we saw in 2.3.3 that the relational analysis required syntactic and semantic complications (*res* movement, structured propositions, infinitely ambiguous *bel*) making the whole thing non-compositional. However, in the current section we’ll uncover major flaws in the Kaplanian analysis, so the relational account eventually beats Kaplan in the category ‘best unified *de re/de se* semantics of belief reports’.

#### 2.3.5.1 Zeevat’s observation

One problematic prediction has been discussed by Schlenker (2003:43,104), who calls it Zeevat’s observation, citing an earlier discussion about it by Zimmermann (1991:206-7). A second, much graver issue is addressed by von
Stechow & Zimmermann (2004). Their point is directly related to the difficulty of proving that a report is false by the Kaplanian semantics, an issue carefully sidestepped (cf. the question marks in Table 2.5) until now. Before launching that more general attack, let’s reproduce Zeevat’s observation about belief in contradictions.

First recap the discussion of Heimson in 1.2.3.2. One of the nice features of Kaplan’s analysis of belief as characterial is that we can account for the cognitive difference between the beliefs expressed by mad Heimson as:

\begin{align*}
\text{(41)} & \quad \text{a. Heimson: “I am Hume.”} \\
& \quad \text{b. Heimson: “Heimson is Hume.”}
\end{align*}

The first is something he often says, for he is indeed insane, but the second is different, Heimson would never say something like that, he is not *logically insane*. Assuming direct reference of proper names to their bearers, and of the first person pronoun to its utterer, we cannot distinguish the two utterances in a merely intensional system, since they express the same proposition, viz. that Heimson is Hume, i.e. $\emptyset$.

Kaplan’s solution to this was the notion of a character, for although both are intensionally contradictory, the sentences do have distinct characters, of which the first is non-empty. In fact, not even full characters are needed for this solution, shifting the notion of a belief alternative from worlds or indices to contexts will do, so the relational analysis of belief can handle it too: Heimson believes *de re* of himself that he is Hume under acquaintance $R = \lambda x \lambda y [x = y]$ but not under $\lambda x \lambda y [y \text{ is (called and known to } x \text{ as) Heimson}]$. So much for Heimson’s belief, now for a report. On Kaplan’s theory the only constraint on truthful reporting is conservation of the proposition expressed. For example:

\begin{align*}
\text{(42)} & \quad \text{Heimson believes he is Hume} \\
& \quad \triangleright \ \text{Bel}_{\text{heimson}}^{\text{that he is Hume}} \\
& \quad [[(42)]^{L_c} = 1 \text{ iff [by (31)] there is a } C \ \delta C \supseteq \mathcal{B} \text{el}^C(\text{heimson, } i) \text{ and } C(\langle \text{heimson, } t_i, w_i \rangle) = [\text{heimson = hume}]^{L_c} (= \emptyset)]
\end{align*}

So, (42) is true if we can find something that Heimson believes and that actually expresses a contradiction. The character of *I am Hume* obviously verifies these criteria, so (42) is true. No problem there, but it is easily seen that the very same character verifies every report about Heimson whose complement expresses a necessary falsehood, so, for example we incorrectly predict that in the Heimson scenario, each of the reports in (43) are true. Moreover, in a sense, they are true by virtue of the fact that Heimson self-
ascribes the contingent/non-empty property of being Hume.

\[
\begin{align*}
\text{(43)} & \quad \text{Heimson believes } \\
& \quad \begin{cases}
\text{Heimson is Hume} \\
\text{he is Napoleon} \\
\text{Napoleon is Hume} \\
0=1
\end{cases}
\end{align*}
\]

In general, it’s a rather immediate consequence of the prohibition of monsters that \(x\) believes that \(\varphi\) and \(x\) believes that \(\psi\) are truth-conditionally equivalent whenever \(\varphi\) and \(\psi\) are (e.g. if both are contradictory).

So we’ve seen a case where the existential definition of belief reports turns out to be too weak. This is partly due to an interaction between direct reference of proper names in possible worlds semantics, and omniscience problems due to belief set closure properties. An important question now is, can part of the reason we incorrectly predict truth in the Zeevat example lie in the fact that it’s just too easy to find verifying characters? Are the two criteria given in (31) enough to appropriately restrict the huge set of all characters, or is the underspecification noted earlier getting out of hand? With this in mind, we may begin to worry about the question marks in table 2.5. We placed these question marks in cases where we wanted to get falsity, and where the obvious candidate characters indeed fail, but where we have not yet proven that no character will do.

### 2.3.5.2 Von Stechow & Zimmerman (2004): the fatal blow

To facilitate the (dis)proving of instances of the existential statement in (31), our reconstruction of Kaplan’s ‘Adding “Says”’ for belief reports, we will present von Stechow & Zimmermann’s (2004) proof of (44), a very interesting theorem of the Kaplanian Logic of Demonstratives with Belief and Reports:

\[
\begin{align*}
(44) & \quad \text{If } [\varphi]_i^c = 1 \text{ then } [\text{bel}^\text{ext}_{\tau} \varphi]_i^c = 1
\end{align*}
\]

A dramatic result with a simple proof based on the massive underspecification of the believed character subsistent in the report semantics of (31).\(^{21}\)

\(^{21}\)This underspecification theorem is very close to an argument by Crimmins (1998:27) who wants to show that “[a] semantic account of [diagonal propositions] seriously underdetermines an account of [horizontal/intensional] content.” The reverse of this claim would relate to the current issue of Kaplanian belief reports, because there the search for a verifying character is constrained only horizontally by the intension of the complement, and diagonally by the belief set of the subject. So let’s translate Crimmins’ (1998:fn.20) argument to our Kaplanian framework:

\[
\begin{align*}
(i) & \quad \text{If } \mathfrak{M} \models^c \varphi \text{ and } \mathfrak{M} \models^c \psi, \text{ then there is } \chi \text{ s.t. } \chi \models \varphi \text{ and } [\chi]^c = [\psi]^c
\end{align*}
\]
Chapter 2. Attitude reports

proof Assume $[[\varphi]]^c_i = 1$. Let $d = \langle \llbracket \tau \rrbracket^c_i, t_i, w_i \rangle \in C$. We build a verifying character $D$ for the report as follows: (i) for all $j \in I$, set $D(d)(j) = [[\varphi]]^c_j$; (ii) for all other context-index pairs $\langle c', j' \rangle$ (i.e. those with $c' \neq d$) set $D(c')(j') = 1$, so that immediately all contexts besides $d$ are trivially in the diagonal of $D$. In fact, even $D(d)(\langle \llbracket \tau \rrbracket^c_i, t_i, w_i \rangle) = [\text{by (i)}] = [[\varphi]]^c_{\langle \llbracket \tau \rrbracket^c_i, t_i, w_i \rangle} = [\text{by definition of } d] = [[\varphi]]^c_{\langle \llbracket \tau \rrbracket^c_i, t_i, w_i \rangle} = [\text{by (i)}] = [\text{ex hypothesi}] = 1$, so $d$ is also in $D$’s diagonal. So, $\delta D = C \supset Bel_C(\llbracket \tau \rrbracket^c_i, i)$. Also, by (i) and definition of $d$, $[[\varphi]]^c = D(\langle \llbracket \tau \rrbracket^c_i, t_i, w_i \rangle)$, so by (31) $[\text{BEL}^\text{that}_{\tau} \varphi]^c_i = 1$. QED.22

Everything that happens to be true is believed by everyone? That’s infinitely worse than classical omniscience, i.e. the prediction that everybody knows all logical truths. Obviously, semantic predictions will be way too weak. Returning to the test cases we’ve been considering throughout the chapter, let’s assume that in all of A, B and C all the belief complements are actually true. This won’t change the intuitions of table 2.1, nor any predictions of earlier theories, especially since the belief complements in question are all in future tense, so their actual truth value should have even less of an effect on the reported beliefs’. Consequently, applying (44), we’d get all 1’s for the question marks in table 2.5, forcing us to discard the Kaplanian semantics of belief reports.

How did this happen? Well, the construction of the verifying $C$ utilizes the underspecification of believed character by the report’s horizontal intension and diagonal belief, worsened by the logical closure intrinsic to the classical use of belief sets in the analysis of the latter. Von Stechow & Zimmermann consider and reject two tentative solutions to the underspecification problem. First, we could give up the reduction of characterial belief to diagonal self-ascription, thus getting rid of the fact that the set of believed characters is closed under ‘diagonal supersets’. Second, we could try adding a restriction on the domain of quantification over characters, in particular,

proof Take $\chi = AN \varphi \land \overline{AN \psi \to \psi}$ and use the following facts provable in Kaplan’s logic (Appendix 1.B, $A =$ actually, $N =$ now): if $\varphi$ and $\psi$ are true in $c$, $[[AN \varphi]]^c = [[AN \psi]]^c = I$; and $\backslash \backslash \backslash AN \ldots \backslash \backslash \backslash \llbracket \chi \rrbracket^c = \backslash \backslash \backslash \ldots \backslash \backslash \backslash$.

Von Stechow and Zimmermann’s proof exploits the same peculiarity of the system: an arbitrary diagonal and horizontal proposition can always be unified in one character as long as they agree in the actual context and actual index (for that’s where they intersect). The added feature here is that it gives a recipe for building a formula expressing the appropriate character. This aspect will be mimicked in the proof of the strengthened version (81a) below.

22In addition to (44), we get from this proof that, if $[[\varphi]]^c_i = 0$ and $\langle \llbracket \tau \rrbracket^c_i, t_i, w_i \rangle \notin Bel_C(\llbracket \tau \rrbracket^c_i, i)$, then too $[\text{BEL}^\text{that}_{\tau} \varphi]^c_i = 1$. 

we’ll consider restrictions to linguistically expressible characters, and to vivid ones.

To start with the discussion about diagonals versus full characters as models of belief. In our reconstruction of Kaplanian belief logic in 1.2.4.3 and 1.B.4 we argued that the cognitive significance of belief resides solely in the character’s diagonal. Thus, we molded Kaplan’s indirect discourse analysis into a diagonal-based belief semantics, with belief sets as sets of belief alternative contexts, \( B_{\text{el}}^C \), by means of the following meaning postulate:

\[
C \in B_{\text{el}}^{\text{Char}}(a, i) \iff \delta C \supseteq B_{\text{el}}^C(a, i) \quad [\approx(74), \text{p. 70}]
\]

Let’s say we give up (45) and take \( B_{\text{el}}^{\text{Char}} \) as a primitive in the definition of belief:

\[
C \in B_{\text{el}}^{\text{Char}}(a, i) \iff \delta C \supseteq B_{\text{el}}^{\text{Char}}(a, i) \quad [\approx(72), \text{p. 69}]
\]

In doing so we get rid of the bug/feature that believing a character automatically implies believing every other character whose diagonal is a superset of the first character’s diagonal. Consequently, we no longer predict that diagonally tautological characters, like the one constructed in the proof, are necessarily believed by the subject. Problem solved? Perhaps, but at what cost? A closer look reveals that this move amounts to giving up an essential distinction between indirect speech (75bi), p. 70, and belief reports (75bii). It thus constitutes a rather drastic shift towards sententialism, severely hampering the logic of belief, as witness the immediate loss of all general principles of rationality, like if \( x \) believes \( \varphi \) and also believes \( \psi \), then \( she \) believes \( \varphi \land \psi \).\(^{23}\) For reasons discussed a bit in 1.1.2 we will shy away from sententialism, preferring a more semantic account with possible worlds or contexts modeling belief alternatives. Besides getting a questionable logic of belief (that may or may not be remedied by independent logical tools) and a less attractive conceptual basis, we lose all other benefits of the reduction to diagonals enumerated in 1.2.4.3. In short, for reasons elaborated upon in that section and more generally in 1.1.2.2, I oppose to this type of sententialism.

\(^{23}\)It may be possible to reintroduce some of these principles to the logic, without falling back into the full logical closure that we started out with, and/or we may resort to other logical tricks, familiar from doxastic logic, to block various kinds of omniscience properties of belief (impossible worlds, awareness logic, compartmentalized minds etc., cf. 1.1.2.3). However, all of this is independent of the issue at hand: back in chapter 1 we argued already that omniscience is a problem, but one that concerns the logic of belief independently, so we were going to refrain from building anti-omniscience machinery into our logic.
ism, maintaining—with von Stechow & Zimmermann (2004) and most other
semanticists working in the tradition started by Kripke, Lewis, Stalnaker,
Montague and others—that, for the purpose of an empirically adequate re-
port semantics, we better stick to the pure possible world/context analysis
and, in this case, the consequent reduction of characters, via diagonals, to
Lewisian properties. Admittedly, a semantic treatment of belief with possible
worlds necessarily results in some logical omniscience, which may have to be
addressed at some point, but as we’d agreed before, not here. Besides, the
problem facing Kaplan goes beyond logical omniscience: it’s a consequence of
the particular underspecification intrinsic to the belief report generalization of
‘Adding “Says”’.

The second attempt at fixing the Kaplanian semantics was to add restric-
tions on the characterial quantification. Still inspired by Kaplan’s analysis of
indirect speech, von Stechow & Zimmermann (2004) propose a restriction in
Kaplan’s belief semantics to characters “expressible in the language spoken
by the subject” (p.13), by which they mean that a believed character must
not only have the diagonal self-ascribed, but it must also be the character of
an (English) sentence that the subject might have uttered or entertained as
a thought.

(47) \( C \in Bel^{\text{Char}}(a, i) \) iff there is a sentence with character \( C \) in \( a \)’s lan-
guage in \( i \) and \( \delta C \supseteq Bel^C(a, i) \)

so \( [\text{Bel}_{\tau}\varphi]_{i^c}^{f} = 1 \) iff there is a sentence with character \( [\varphi]_{i}^{f} \) in
\( [\tau]_{i^c}^{f} \)’s language in \( i \) and \( \{\varphi\}_{i^c}^{f} \supseteq Bel^C([\tau]_{i^c}^{f}, i) \)

Using that, we get the following for reports:

(48) \( [\text{Bel}_{\tau}\varphi]_{i^c}^{f} = 1 \) iff there is a sentence with character \( C \) in
\( [\tau]_{i^c}^{f} \)’s language in \( i \) and \( \delta C \supseteq Bel^C([\tau]_{i^c}^{f}, i) \) and \( [\varphi]_{i^c}^{f} = C([\tau]_{i^c}^{f}, t, w) \)

This, again, brings a form of sententialism into the belief semantics, but
perhaps an acceptable one, at least by von Stechow’s and Zimmermann’s
standards, so let’s try it on.\(^{24}\) It was hoped that this would block the very

\(^{24}\) The argument for diagonals, against characters, in chapter 1 only applies to belief,
so the expressibility condition will definitely be acceptable if we leave it outside the logic
of belief and just put it in the report semantics, i.e.:

(i) a. \( [\text{Bel}_{\tau}\varphi]_{i^c}^{f} = 1 \) iff \( \{\varphi\}_{i^c}^{f} \supseteq Bel^C([\tau]_{i^c}^{f}, i) \)

b. \( [\text{Bel}_{\tau}\varphi]_{i^c}^{f} = 1 \) iff there is a sentence in \( [\tau]_{i^c}^{f} \)’s language at \( i \) with character
\( C \) and \( \delta C \supseteq Bel^C([\tau]_{i^c}^{f}, i) \) and \( [\varphi]_{i^c}^{f} = C([\tau]_{i^c}^{f}, t, w) \)
2.3 De se and de re

artificially constructed, abstract, near tautological character that was used as witness in the proof of (44). Indeed the restriction seems quite heavy, and in fact it excludes most characters, because our formal (and natural) languages are but countably infinite so there are only countably many expressible characters, whereas time is often assumed to be uncountable (modeled as real numbers in physics), so the set of contexts and thus the set of characters must also be uncountable.

However, as pointed out by von Stechow & Zimmermann (2004) crediting Philippe Schlenker (p.c.), a restriction to expressible characters is not enough. We can still prove the following special case of (44):

\[(49) \text{If } M \models_{c} \text{win}(a) \text{ and } \text{win}(a) \lor \neg \text{ANwin}(a) \text{ is expressible in } \llbracket e \rrbracket_{(w_{c}, t_{c})}^c \text{'s language in } c, \text{ then } M \models_{c} \text{BE}_{\text{tharwin}}(a)\]

**proof** The proof consists in showing that \(\llbracket \text{win}(a) \lor \neg \text{ANwin}(a) \rrbracket\) is a verifying character for the belief report. That it is then also an expressible character follows from the premises, which are easily fulfilled, given that the logical form of *Either Arata will win, or it's not actually (now) the case that she will*, for many English speakers, is \(\text{win}(a) \lor \neg \text{ANwin}(a)\).

So, assume \(M \models_{c} \text{win}(a)\), and let \(D = \llbracket \text{win}(a) \lor \neg \text{ANwin}(a) \rrbracket\) and \(d = \langle \text{ellsworth}, t_{c}, w_{c} \rangle\). With expressibility out of the way, we check the other two criteria:

- **\(\delta D = \{ c' \in C \mid \llbracket \text{win}(a) \lor \neg \text{ANwin}(a) \rrbracket_{(w_{c}, t_{c})}^{c'} = 1 \} = \{ c' \in C \mid \llbracket \text{win}(a) \rrbracket_{(w_{c}, t_{c})}^{c'} = 1 \text{ or } [\text{ANwin}(a)]_{(w_{c}, t_{c})}^{c'} = 0 \} = \{ c' \in C \mid \llbracket \text{win}(a) \rrbracket_{(w_{c}, t_{c})}^{c'} = 1 \text{ or } \llbracket \text{ANwin}(a) \rrbracket_{(w_{c}, t_{c})}^{c'} = 0 \} = C \supset \text{Bel}_{c}(\langle \text{ellsworth}, (w_{c}, t_{c}) \rangle);**
- **\(D(d) = D(\langle \text{ellsworth}, t_{c}, w_{c} \rangle) = \llbracket \text{win}(a) \lor \neg \text{ANwin}(a) \rrbracket(\langle \text{ellsworth}, t_{c}, w_{c} \rangle) = [\text{no person indexicals}] = \llbracket \text{win}(a) \lor \neg \text{ANwin}(a) \rrbracket(c) = \{ j \in I \mid [\text{win}(a)]_{j} = 1 \text{ or } [\text{ANwin}(a)]_{j} = 0 \} \)**

Otherwise, as we take it in (47), i.e. construed as a principle of belief, it may still make sense, but then it cannot be a merely characterial restriction, i.e. in order to have any effect, it must constitute a restriction on diagonals, since belief was argued in 1.2.4.3 to depend only on diagonals. From the cardinality argument above (there are more characters than formulas and sentences) and the fact that there are as many diagonals as full characters, it follows that the expressibility restriction in the semantics proposed in (47) is so strong that it will indeed restrict diagonals. Even if this is a big step towards sententialism, it’s still defensible as it doesn’t mean we have to rely on non-diagonal points of believed characters.
\[ \{ j \in I \mid [\text{win}(a)]_c^j = 1 \} \cup \{ j \in I \mid [\text{win}(a)]_{(w,c,t)}^j = 0 \} \]

So, \([\text{bel}^\text{tan}\text{win}(a)]_{(w,c,t)}^c\) = 1 QED

In general, Schlenker’s argument shows that any sentence without person indexicals that is part of someone’s language and that is actually true, is therefore believed by that subject. In particular, it helps us find expressible verifiers for all the question marked positions in table 2.5. In other words, even with the significant restriction of expressibility, we cannot exclude enough unwanted verifying characters to save the Kaplanian analysis of belief.

Finally, von Stechow & Zimmermann (2004) mention a suggestion of Uli Sauerland’s (p.c.) to exclude characters like those constructed according to the recipe in the previous proof, on grounds of their lack of vividness à la Kaplan (1969). To counter this last attempt to repair Kaplanian report semantics, let us briefly consider what a vividness restriction in belief would amount to.

First note that, for Kaplan and the later development of vivid names into acquaintance relations, vividness was really a relation between the belief’s subject and a res. If we want to use that, we need something like structured propositions and res movement to keep the dictum and the res apart and we end up with something quite like the relational analysis, with all distinguishing non-compositional aspects. Evidently, a different type of vividness is intended.

If vividness is to work within Kaplan’s compositional framework, it should apply to whole sentence characters: \([\text{bel}^\text{tan}\varphi]_{i}^{f,c} = 1 \text{ iff there is a vivid character } C \text{ s.t. } \delta C \supseteq \ldots \). But a character is not vivid per se, it can only be vivid for a certain individual at a certain index. Among the characters, or rather, diagonals, that an individual believes, there are certain vivid ones, and we can take only those as verifiers. Presumably, Schlenker’s \([\text{win}(a) \lor \neg \text{an}\text{win}(a)]\) is not vivid here. Since the vividness restriction is built into belief, it’s really only the diagonal that counts, so, it follows that in this case, the tautology is not vivid. Apparently, the vividly believed propositions form a subset that is not closed under logical consequence/superset. It seems we have slowly drifted into something akin to awareness logic (cf. discussion in 1.1.2.3), combating the underspecification problem with quite heavy anti-omniscience machinery, which we intended to steer clear of. Besides a commitment to a conceptually dubious and often criticized split between explicit and implicit belief, we might ask if a mere set of contexts is enough information to decide whether or not it constitutes a vivid diagonal. Or should we give
up our adherence to diagonals and go sentential, thus combining a couple of
the strategies discussed above, as in true reports require a character that’s
expressible with a sentence vivid in the subject’s mind or something?

Instead of resorting to sententialism, I would sooner try to go contextual,
by analogy with the suggestions in 2.3.2: A report is true in a context iff
that context provides a (vivid, expressible) character $C$ s.t. . . . . It seems not
quite reasonable to assume that contexts like our test cases A, B, and C
do not provide as salient characters the near-tautological concoctions of the
two above proofs, thus blocking all the unwanted predictions. On the other
hand, it remains to be made precise how such contexts sometimes can provide
verifying characters for the points where we do want them (cf. the 1’s in
table 2.5). And here, the differences between the relational and characterial
analyses of attitudes, as described back in 1.2.4.4, become crucial. Remember
that the main theoretical difference was pinpointed in the analysis of the wide,
‘in the world’ aspect of content: in the relational account the res and the
relation of acquaintance linking the believer and res are to be found in the
actual world; in Kaplan’s theory it’s a sentential character that is sought.
Moving to report semantics this meant we were able to contextualize the
relational account by shifting the wide part, there is an $R$ s.t. $R(a, b)$ and . . . ,
into the context: the context provides an $R$ with $R(a, b)$, and . . . . What this
means is that, prior to an utterance of a de re report, a suitable relation
of acquaintance has to have been established in the context ($\approx$ common
ground). Thus given a res and acquaintance relation, the new information
carried by the report merely consists in the assertion that $a$ believes that it
(the res as represented to $a$ via $R$) has the property predicated in the report’s
complement clause.

Now see what happens if we try and apply the same trick of contextualiza-
tion to Kaplan: there is a character $C$ s.t. . . . becomes the context provides
a character $C$ s.t. . . . , meaning that by the time the report gets interpreted
by the hearer the characterial mode of presentation of the reported belief
should already be old news, just as res and acquaintance are assumed given
in the contextualized relational analysis. But then what new information can
the report’s utterance bring? Given the characterial mode of presentation
($C$), the speaker ($a$), and the evaluation parameters ($c$ and $i$), the report
cannot tell us anything we don’t already know, for in that case the comple-
ment’s content is already fully recoverable (cf. $[[\varphi]]^f_c = C((a, t_i, w_i))$ from
(31)). Note that it’s exactly the res movement separating believed predicate
and res and thereby making the relational account non-compositional, that is
here turned into a virtue by allowing a sensible contextualization. Thus, I’m
even tempted to say that it’s precisely Kaplan’s compositional interpretation
of report complements that causes its downfall. From now on, we’ll focus on
the relational analysis and take that as the starting point for a truly dynamic account of belief reports in chapter 3

2.4 Anti-reductionism and pure de se separatism

Over the previous sections we have reviewed and tested some accounts of belief reports, covering various parts of the de dicto - de re - de se spectrum. For the first half of our test schema in table 2.1, de dicto and de re, we were able to match our semantic judgments quite accurately with simple scope, or with Kaplan’s (1969) reduction of de re as de dicto under a vivid name. With Cresswell & von Stechow’s (1982) Lewis-inspired generalization, we extended that to cover the whole de re/de se area. Doing so, we got a nice unified account of de re/de se in accordance with the arguments in chapter 1 that established that de se is but a special case of de re. An immediate bonus of the unification is the account of Arata believes that she will win as compatible with various de re scenarios, from pure de se (Arata: “I am going to win”) to mere de re (Arata (with amnesia): “Arata will win”). The theory of relational attitudes achieves this by stipulating that English reports are underspecified for mode of presentation of the reported belief. The same is true for Kaplan’s compositional, but on closer examination fundamentally flawed, two-dimensional analysis of belief reports. There, the so-called prohibition of monsters implies a necessary underspecification of reports in natural language. Because of this underspecification, both analyses fail with respect to (1c), a report that is only true for a first person de se belief.

In this section we focus on the remaining problems in the de re/de se domain. In particular, we review the arguments for and against the unification of de re/de se. We start with Chierchia (1989), who argues for a semantic ambiguity between de re and pure de se logical forms, on the basis of the infinitive and gerund data that the relational and non-monstrous character analyses can’t handle. In reaction, Zimmermann (1991) analyses the same pure de se data by means of a different, less pervasive ambiguity, thus creating a kind of hybrid theory, with a Chierchia-inspired separate IIf for infinitival complement constructions, while following Kaplan (1989) (and the relational analysis) in not admitting an ambiguity in the embedded coreferential pronoun cases. Meanwhile, Chierchia’s other arguments for the ambiguity, based on abstract anaphora and elided (ambiguous) belief reports, are attacked by Reinhart (1990), who argues directly against Chierchia and in favor of a variant of the relational analysis. The data, arguments and analyses originating
in Chierchia’s work will be discussed in 2.4.1.

In the realm of quantified belief reports there have also been arguments for and against de se separatism. In 2.4.2 we first present Zimmermann’s observation about everybody believes that . . . , which points in the direction of a unified account like the relational analysis. On the other hand, Percus & Sauerland’s (2003a) arguably quantificational only x believes that . . . datum is meant to support Chierchia’s ambiguity thesis. In this connection we also reheat the discussion about contextualism of acquaintance, as addressed already in 2.3.2.1, because whether and how acquaintance relations are themselves quantified over at If or contextually given, will affect the interpretation of these explicitly quantified data.

Meanwhile, Schlenker (1999; 2003) launches yet another attack on the classical theories by first adding to Chierchia’s infinitives and gerunds more examples of reports that retain a certain pure de se perspective from the original belief (e.g. logophoric pronouns in Ewe and Bafut). In addition, he discusses examples of shifted indexicals in belief reports and (free) indirect discourse. Shifted indexicals are indexicals that directly violate Kaplan’s (1989) Principle 2 and thus his prohibition of monsters, in other words, their reference is not fixed by the actual speech act but shifted by a (monstrous) operator. As Schlenker points out, these monstrously shifted indexicals occur systematically in attitude and speech reports in some languages. He shows for example that the first person pronoun of Amharic, which is just as indexical as English ‘I’, can be interpreted as shifted by attitude verbs, something explicitly ruled out by Kaplanian and relational analyses:

(50) John says that I am a hero [Pseudo-Amharic, (Schlenker 2003)]
    ‘John says that he’s a hero’

Schlenker proposes an account that treats all attitude reporting operators as quantifiers over contexts, i.e. monsters. This leaves a lot of work to be done to derive the old-fashioned Kaplanian English cases. In 2.4.3 we discuss Schlenker’s system(s) and the relationships with our earlier accounts of de re/de se reporting.

Schlenker’s rejection of the Kaplanian analysis (and consequently its sister, the relational account) has stirred up quite some discussion and this time it’s von Stechow (2002) who steps up to defend (what he calls) the classical theory, employing a mechanism of feature deletion to enforce bound-variable interpretations for embedded pronouns and other (apparent) indexicals. In 2.4.3.5 we set out to compare this reaction with its direct rival and the earlier accounts.
2.4.1 Pure de se reports

As we have seen, both unified accounts of de re/de se predict that reports do not and cannot convey purely indexical distinctions. In other words, the exact mode of presentation (be it acquaintance relation or character) of the original thought gets lost upon reporting. This section discusses some constructions in English that contradict that prediction and thus undermine the unified accounts. In each case, we arguably have a belief or other attitude reporting construction that unambiguously reports a pure de se attitude. We also look at the alternative theories that have been proposed in light of these data.

2.4.1.1 Indirect reflexives and he*

The claim that English belief reports can distinguish between general de re and pure de se can be traced back at least to Geach (1957) who considers the contrast between:

(51)  

a. Philip […] believes that Philip’s worst enemy is dead  
b. Philip […] believes that he himself is dead [(Geach 1957:129)]

Geach remarks that these are not equivalent, because unlike (51b), (51a) may have a true reading, for instance if Philip is his own worst enemy though he does not realize that.\footnote{For reasons that do not concern us here, Geach’s original examples had the more complex (yet in the crucial case coreferential) Philip’s worst enemy as matrix subject.} He analyses (51b) as “containing what classical grammarians call an indirect reflexive pronoun”, viz. he himself, which serves as “an oratio obliqua proxy for the first-person pronoun of oratio recta” (Geach 1957:129). Apparently, according to Geach’s intuitions, reports with reflexives as in (51b) are only true if the reported thought is first person, in this case the absurd “I am dead.” This observation contradicts our unified analyses of de re/de se that predict underspecification with regard to the mode of presentation for all reports.

Castañeda (1966) takes up the study of de re/de se reporting, arguing for an enrichment of the logical language used in semantics to capture the distinction between pure de se and general de re reports. In particular, he introduces the term he* that does exactly what Geach claimed about he himself, so:\footnote{The distinction is independent of a de dicto/de re ambiguity in (51a), because even on the de re reading, with singular proposition as complement, the fact remains that it has this true reading of mistaken self-identity. The existence of a dedicated pure de se pronoun gives rise to the question whether there are any languages that grammaticalize intentional de se reports? Or just non-first...}
2.4 Anti-reductionism and pure de se separatism

(52) a. \[\text{[Avis thought that she was looking good]} = 1\] iff Avis expressed the proposition that Avis is looking good, e.g. in cases like:
   - Avis: “I look good”
   - Avis [pointing to photo of herself]: “She’s pretty”
   - Avis, with amnesia: “Avis Fish is looking good”

b. \[\text{[Avis thought that she}^*\text{ was looking good]} = 1\] iff Avis: “I look good”

Castañeda calls the starred, pure de se reporting pronoun of (52b) a quasi-indicator (other names sometimes used to capture this semantic property include de se pronouns or logophoric pronouns, but we’ll discuss that last term in depth in 2.4.1.2). By definition, quasi-indicators convey the first person perspective in a report and therefore the classic theories of unified de re/de se will deny that he\(^*\) has a counterpart in natural language. However, Castañeda himself seems to suggest that English (he) himself when it occurs as subject of a report complement, is just that, a quasi-indicator, which is exactly what Geach (1957) had claimed before him. It’s debatable whether he himself has a truly distinct indirect reflexive (aka long distance reflexive) person de se? Or de hic, or even de re non-de se? Given the division of pragmatic labor (marked forms ↔ marked meanings), you might even expect there to be a he\(^*\) for:

(i) \[\text{[Avis thought that she}^*\text{ was looking good]} = 1\] iff Avis expressed the proposition that Avis is looking good, without referring to herself in the first person, e.g. by thinking: “She’s pretty” while looking at a photo of herself.

For it’s obvious that the pure de se reading is the unmarked meaning of English coreferential reports, i.e. of the form NP, believes that NP, VP. Wouldn’t it be cheaper to have a special construction dedicated to the marked meaning? I guess the marginality of coreferential impure reports (mistaken self-identity) would provide a negative answer there. But still, assuming Castañeda and Geach are right about the existence of he\(^*\) and not he\(^*\) in the (English) lexicon, we’d expect a pragmatic shift in meaning for the general de re/de se pronoun to take on the leftover meaning, i.e. impure de re/de se (compare it with the classic lexical pragmatics example of cause to die taking on the marked, indirect meaning in view of the lighter alternative kill, though I can’t really say which of he and he\(^*\) should be considered the lighter form). In English this pragmatic prediction is not borne out at all, there’s not the slightest hint of impure de se if a reporter just uses a regular coreferential pronoun instead of a he\(^*\) or he himself, unless the context very explicitly sets it up (and even then its reportedly hard to get). However, in some African languages with so-called logophoric pronouns (≈ he\(^*\), see below in 2.4.1.2), it has been noted that the non-logophoric third person pronominal forms take on an anti-logophoric meaning (≈ everything but pure de se) (Schlenker p.c.). Unfortunately, but for Kusumoto’s (55b) below, the typological literature shows only anti-logophoric pronouns not interpreted coreferentially with the matrix subject at all, so more empirical evidence is needed.
use, and if so, whether indirect reflexives (including the ones in Ancient Greek and Latin) really are quasi-indicators, semantically.

Schlenker (2003:59), citing a discussion by Clements (1975), notes that the classical reflexives and *he himself* differ at least in one important respect from *he*, viz. their distribution: English *he himself*, Latin *se*, and Greek *heautos* also occur as so-called direct reflexives, i.e. bound by c-commanding antecedent, whereas *he*, by definition, grammaticalizes the role of quasi-indicator unambiguously and can therefore never occur outside an indirect discourse complement. The question remains whether, assuming the direct and indirect uses are really distinct and separable, the indirect one is a quasi-indicator. Also, it seems that Anscombe’s (1975) remark about Ancient Greek cannot be so easily discarded, since the rare forms she quotes as natural language instantiations of *he*, viz. *hoi, he* and *how*, are only allowed in report complements, so there it appears there are dedicated indirect reflexives.

To settle the matter of whether indirect reflexives are really quasi-indicators we need to look at reports of mistaken self-identity with mirrors or the like. For English *he himself* we can ask native speakers, but unfortunately, judgments are not clear. On the side of Geach and Casta˜neda (a nonnative English speaker) we have, for instance, Sells (1987), Chierchia (1989), Reinhart (1990) and Atlas (p.c., to appear). To maintain such a position we need to distinguish the direct and indirect reflexive *he himself* from a third, focus related usage of *he himself* to signal contrast or emphasis. Atlas, for instance, claims that a phonetically unstressed *himself* in American English is truly quasi-indexical. The data appear to be something like:

(53) Looking at a photo in a portfolio, Avis Fish, a really bitchy America’s Next Topmodel judge, says: “Who’s that? She looks awful”. What she didn’t know was that she was actually pointing at a rather unflattering picture of herself, so it seems that, despite her apparent self-confidence Avis actually believed that she or she herself \[\{\text{looked} \]
\[\{\text{she, she HERSELF} \]
\[\{?\text{she herself, *she*}\] looked awful.

For the classical indirect reflexives it’s even harder to find hard evidence. Descriptive grammars are inconclusive but the literature contains some suggestive evidence for the claim that indirect reflexives are quasi-indexicals. Consider for instance the following quote from a study about reflexives in the New Testament, immediately following an almost complete list of the

\[\text{http://www.upn.com/shows/top_model/}\]
indirectly reflexive occurrences of _heautos_: 

What distinguishes these examples from others where the reflexive is not used in the embedded clause is that these verbs all indicate either the thought or the expression of the subject. Such verbs may be called “logophoric.” We may therefore provisionally propose that when a complement clause of a logophoric verb contains a pronoun which is co-referent with the subject of the logophoric verb, the pronoun will be reflexive even if it is not co-referent with anything in its own clause. Such pronouns may be also called “logophoric.”

[(Tiller 2001:53)]

In other words, indirect reflexives are logophors, which means they only occur in reports that represent the subject’s thought or words. After looking at some remaining examples the definition of logophoricity is made more precise by, among other things, the following:

A logophoric verb is a verb of saying or thinking that takes a complement clause which contains the thought or expression of the subject. […] a logophoric verb must portray the thought as consciously present to the mind of the thinker. [(Tiller 2001:55)]

Does the belief complement in (53), _that she [herself] looked awful_ represent “the thought or expression of the subject […] as consciously present to the mind of the thinker”, i.e. Avis? No; Avis is brimming with self-confidence and the thought of looking awful herself would never cross her mind. If this is correct, the Greek indirect uses of reflexives are indeed quasi-indexicals, which concludes a demonstration of the inadequacy of _de re/de se_ unifications that predict a total underspecification of the original thought’s mode of presentation.

2.4.1.2 Logophors

In the previous subsection we introduced the concept of logophoricity in relation to indirect reflexives. It was claimed that indirect reflexives are logophoric and logophoricity was defined, a bit vaguely, as reporting the conscious thought or utterance of a subject. There seems to be a connection between the notions of logophoricity and quasi-indexicality, which had gone unnoticed until Schlenker (1999) (crediting Heim’s 1991 classnotes) equated them. To reach his conclusion, that logophoric pronouns are natural language instantiations of _he^⋆_, Schlenker uses the definition of logophoricity by
Clements (1975) with data from Ewe\textsuperscript{29}, Gokana\textsuperscript{30} and Bafut\textsuperscript{31}.

A typical example of logophoric vs. anaphoric/deictic reference:

(54)  
\hspace{1cm} a. kof\text{\small \textit{i}} be \textit{y`e} -dzo  
\hspace{1cm} \textit{Kofi say LOG -leave}  
\hspace{1cm} \textit{‘Kofi, said that he\textsubscript{i}/\textsubscript{j} left’} 
\hspace{1cm} 
\hspace{1cm} b. kof\text{\small \textit{i}} be \textit{e} -dzo  
\hspace{1cm} \textit{Kofi say 3.sg -leave}  
\hspace{1cm} \textit{‘Kofi, said that he\textsubscript{j} left’}\textsuperscript{32}  

\hspace{1cm} \textsuperscript{[Ewe (Clements 1975:142)]}  

Apparently, the logophoric pronoun (LOG) is used to signal co-reference between matrix and embedded subjects. Additionally, logophors only occur in report contexts, and must reflect the reported thought from the point of view of the thought’s subject. These three criteria, according to Schlenker (2003:59), make logophoricity “remarkably similar” to quasi-indexicality as defined by Casta\~neda. The decisive data, infelicity of LOG in mistaken self-identity reports, were provided later by Kusumoto (1998):

(55) John is looking at a mirror from a distance and sees a man in the mirror. He notices that the man’s pants are on fire. In fact, the man he sees in the mirror is John himself, but he doesn’t realize it.

\hspace{1cm} a. John believes that his pants are on fire 
\hspace{1cm} b. John wà?át\textsuperscript{\textalpha} m\textbeta *yu/\textalpha k\textbeta kí  
\hspace{1cm} \textit{John thinks that *LOG/he FUT burn}  
\hspace{1cm} \textit{‘John thinks that he’s going to get burnt’}  

\hspace{1cm} \textsuperscript{[Bafut (Kusumoto 1998)]}  

It seems clear then that logophors really are the natural language counterparts of he\textsuperscript{*}, matching in semantic (co-referent with matrix subject, reporting pure de se thoughts of that subject), syntactic (occur only in report contexts), and morphological properties (3rd person). Their peculiar semantic property of unambiguously reporting the thought as perceived from the point of view of the subject, i.e. purely de se, was a crucial part of the evidence that led Schlenker to denounce Kaplan’s analysis and replace it with a system that treats attitude operators as genuine monsters.

---

\textsuperscript{29}A language spoken in Ghana and Togo by approximately three million people  
\textsuperscript{30}A language spoken in Nigeria  
\textsuperscript{31}A Bantu language of Cameroon  
\textsuperscript{32}Usually, in such examples, j \neq i, but this need not always be the case, as (55b) shows. On the other hand, some schools view these subscripted indices merely as indicators of syntactic binding, in which case you’d probably say that the indices here are always distinct, but (pragmasemantically) coreference might still occur (as accidental coreference, (Lasnik 1976)).
However, let me point out some qualifications to Schlenker’s claim. First, as for logophors only occurring in attitude or speech reports, consider:

\[(56)\]  
\[\text{dëvi} -\text{a} \, \text{xọ} \, \text{tohehe} \, \text{be} \, \text{yè} -\text{a} \, -\text{ga} -\text{da} \, \text{alakpa} \, \text{ake}\]  
\[\text{child} \, -\text{D receive punishment so that LOG -T -P} \, -\text{tell lie \, again}\]

\[\text{o} \, \text{NEG}\]

‘the child, received punishment so that he, wouldn’t tell lies again.’  
\[[\text{Ewe (Clements 1975:160-1)}]\]

According to Clements the logophor here suggests that “the child voluntarily received punishment, in the belief that this would cure him of his untruthfulness” (p.161), i.e. it indicates that the second clause, describing the purpose of the punishment, is to be understood from the child’s point of view. So, in a way, the logophor does represent the antecedent’s subjective point of view here, but there is no indirect discourse embedding to be found, contradicting many a theory of logophoricity.

Second, the first person point of view criterion, exemplified by (54) and (55b) and perhaps still satisfied on a vague, implicit level in (56), is too strict. This is shown in (57) where the reported speech seems to have been second rather than first person:

\[(57)\]  
\[\text{âmâ} \, \text{lamá} \, \text{gospel} \, \text{kọ} \, \text{m} \, -\text{ê} \, -\text{dọ} \, -\text{ê}\]  
\[\text{I \, tell \, Gospel \, COMP \, COMP \, -3 \, -fall \, -LOG}\]

‘I told Gospel, that he, fell.’  
\[[\text{Eleme (Bond 2004:7)}]\]

Obviously, this is a reports of the matrix subject saying: “You fell.” Besides reporting a non-first person thought, it is (therefore) also a report about the reporter’s own speech, not Gospel’s, who is the antecedents of the logophor.

Let me immediately point out that, interestingly, the data are not uniform with respect to allowing second person speech to be reported with logophors, even among closely related languages:\[33\]

\[(58)\]  
\[*\text{mùn} \, \text{kọ} \, \text{nê} \, \text{lèbàrè} \, \text{kọ} \, \text{aè} \, \text{dọ} \, -\text{ê}\]  
\[\text{I \, said \, give \, Lebare \, that \, he, \, fell} \, -\text{LOG}\]

‘I said to Lebare, that he, fell.’  
\[[\text{Gokana (Hyman & Comrie 1981:22)}]\]

Apparently, the standard definitions of logophoricity or quasi-indexicality cannot quite capture the full range of data: In some languages it may be the case that third person (accidental) coreference, as in mistaken self-identity scenarios (cf. (55)), cannot be reported logophorically, while second person

\[33\text{Both Eleme, (57), and Gokana are so-called Ogonoid languages spoken in Nigeria.}\]
intentionally disjoint reference can (cf. (56), (57)). In the next subsection we will see that the English construction dedicated to reporting pure de se belief also allows for second person de se in addition to just first.\footnote{Chierchia (1989:17) and Schlenker (2003:61) give examples like:}

Picking up on these second person de se reports, Culy (1994; 1997) and Bond (2004) argue against the idea that logophors convey any point of view at all. These authors propose that the “[p]rimary function of morphologically distinct logophoric pronouns is to mark discourse as ‘reported’ and not to reflect point of view” (Bond (2004:1) paraphrasing Culy (1997)), which unfortunately lands us right back in the first problem, the one posed by (56), containing no indirect discourse. More needs to be said, evidently, about the relation between he*, logophors, indirect reflexives, and 1st and 2nd person de se reporting, but we’ll have to leave that for future research.

\subsection*{2.4.1.3 Chierchia’s PRO and the ambiguity thesis}

In 2.3.1 we spent some time establishing that—for some attitude verbs in some languages—there is a genuine semantic contrast between coreferential 3rd person pronoun reports and reports with subjectless infinitives and gerunds as complements. Examples considered so far to make this point include, first, the contrast between 2.1’s (1c) and (3) in the mistaken self-identity scenario B:

\begin{itemize}
  \item[(i)] John [to Mary, whom he doesn’t recognize]: “Mary should leave”
    \begin{itemize}
      \item a. #John told Mary to leave.
      \item b. John told Mary that she should leave.
    \end{itemize}
\end{itemize}

This also shows that it is sometimes possible to distinguish general de re reports, compatible with mistaken identity, from a specific subclass of de se, but this special subclass apparently includes not only pure first person attitudes, but also (in some cases) second person speech/attitudes. As with the African logophors, there’s variation across languages and also across different attitudes within a language, deserving of further research. Further, connecting this discussion of 2nd person de se with logophors and quasi-indexicals, another interesting question comes to mind: What happens if we introduce a Geachian he himself as embedded subject of a command report?

\begin{itemize}
  \item[(ii)] ??John told Mary that she herself should leave.
\end{itemize}

I have no data about this last one, but it seems out, unless perhaps the reflexive emphatically marks some kind of contrast, thus reporting e.g. “MARY HERSELF should leave!”? Recall the discussion about Atlas and the uses of reflexives in 2.4.1.1 where it’s suggested that the emphatic use of the reflexive is distinct from the logophoric. In fact, I’m even unsure about the status of (ii) in a non-pathological 2nd person context, where the reported command was “YOU YOURSELF should leave.”
(59) Ellsworth, hearing people talk about Arata Suggs while watching her speech on TV, says: “That girl, Arata, I’m sure she’ll win” In fact this Arata Suggs is none other than herself, she just didn’t recognize the image on the TV, nor did she know of this nickname the others use to refer to her. 
   a. Ellsworth believes she’ll win. 
   b. #Ellsworth believes to be on the winning side.

The same point was made in 2.3.1 with the attitude verb hope:

(60) Thinking she’s seeing her opponent, but actually seeing herself, Essie mutters: “I hope you lose.”
   a. Essie hopes that she’ll lose. 
   b. #Essie hopes to lose.

which was strengthened in the form of a contextless infelicity/ungrammaticality judgment:

(61) Although she doesn’t realize it herself, Avis believes that she is
    *believes to have been
    hopes that she is
    *hopes to be
    elected.

Chierchia (1989) uses such judgments—in Italian—as the empirical basis for his argumentation against unified de re/de se semantics, which he identifies with Boër & Lycan’s (1980) propositional unification, cf. 1.2.4.1. The data show that, contra predictions of not only the arguably flawed account of Boër & Lycan (1980) but also more sophisticated unifications by Cresswell & von Stechow (1982) and Kaplan (1989), there is a linguistically real distinction between pure de se and other de re, even in English belief reporting. The contrasts show that reports with infinitival complements, like (59b), are unambiguously pure de se, whereas the pronoun report (59a) is compatible with any de re or de se mode of presentation.35

35Following up on footnote 34, recall that Chierchia’s data include infinitival command reports that are unambiguously 2nd rather than 1st person de se:

(i) John to Mary, whom he doesn’t recognize: “Mary should leave”    [= (i), p. 118]
   a. John told Mary that she should leave. 
   b. #John told Mary to leave.

If we simply restrict attention to belief, we do not have to modify the generalization that PRO correlates with a first person perspective, but in the interest of extendability these
Chapter 2. Attitude reports

Chierchia’s main thesis is what I shall refer to as the *ambiguity thesis*: *de re/de se* reporting does not form a homogeneous class. Within it, we need to distinguish two logical forms, one for pure, first person, *de se*, and one for *de re*. Chierchia cashes out this distinction semantically in terms of the distinction between property self-ascription vs. propositional belief. The first person *de se* readings (which he calls just *de se*) correspond to Lewis-style property self-ascription:

\[ \text{Bel}^*_x \lambda x[\text{win}(x)] \]

\[ \text{[}(62a)\text{]}_w = 1 \quad \text{iff} \quad \text{Ellsworth self-ascribes the property of winning} \]

data should be taken into account at some point. The contrast in (i) itself does not really pose a problem, what makes the issue more puzzling is the addition of Schlenker's (p.c.) scenario, which seems to turn the judgments around:

(ii) John, dividing part of his inheritance over his as yet unborn third generation descendants: “My great-grandchildren should behave, otherwise…”

   a. ...in his will John told his great-grandchildren that they should behave
   b. ...in his will John told his great-grandchildren to behave

What is the relation of acquaintance between John and his unborn great-grandchildren? Since he doesn't know them yet, how can the command expressed in the will be *de re*? Let alone 2nd person *de se*, because, for the same reason, the direct discourse version, the written text of the will, cannot have contained any (irreducibly 2nd person) imperatives or *you’s* to talk about those great-grandchildren. Note that this is not so much a problem for unificatory analyses that heap *de re* and *de se* together, but rather for 1st+2nd person *de se* separatist proposals like Chierchia’s ambiguity thesis below, which predicts the reverse of Schlenker’s judgment (as in (i)). A unificationist way out is to argue that actually both are OK, which is what she’d predict, since they both report a general *de re/de se* belief under a somewhat indirect (*de dicto*-ish) relation of acquaintance. That sentences like (iiib) are quite OK is corroborated by some similar, Googled examples:

(iii) a. God told his people that they should multiply, he said “Be fruitful and increase, fill the earth” (Genesis 1:28).

   b. When he got back to Kapiti [from his last raid on the South Island] he told his people that they should go to Otakou and Rakiura to kill the men of Te ...

   c. President Daniel arap Moi of Kenya told his compatriots that they should refrain from sex for two years to stop the spread of AIDS.

Is God really acquainted with “his people”? Isn’t that just as indirect as John’s relation to his grandchildren via his will? Or presidents with their people? The data need further attention, but perhaps an important observation is that all of these examples, including Schlenker’s, seem to involve *de dicto* reporting. In addressing their audience, the subjects may or may not have used 2nd person pronouns, but they intended their command to be *de dicto*, i.e. *his great-grandchildren, whoever/wherever/whatever/whenever they might turn out to be; my people as in whoever chooses to follow me*, etc.
\[
\text{iff } \left[ \lambda x[\text{win}(x)] \right] = \mathbb{I}(\text{win}) \supseteq \text{Bel}(\text{ellsworth}, w)
\]

This, then, is the LF of the infinitival report (59b), which accounts for the falsity in the 3rd person mistaken identity scenario B. The pronoun report (59a) has this same reading most of the time, but, says Chierchia, this type of report is actually ambiguous between that and a de re reading represented by the following LF:

(63)  
\[ \begin{align*}
a. & \exists x[x = e \wedge \text{BEL}_x \text{win}(x)]^{36} \\
b. & \left[ (63a) \right]_{w} = 1 \\
\text{iff } & \text{Ellsworth believes de re of Ellsworth that she is winning}
\end{align*} \]

The BEL operator has to defined in such a way the whole de re LF is true in B, for which purpose any variant of the de-re-under-vivid-description framework, in the line of Kaplan-Lewis-Creswell-von Stechow, would do fine. For instance, we can opt for a de re LF à la Kaplan (1969):

(64)  
\[ \begin{align*}
a. & \exists d[R(d, e, e) \wedge \text{BEL}_e \text{win}(\gamma d)] \\
b. & \left[ (64a) \right]_{w} = 1 \\
\text{iff } & \text{there is a vivid name } d \text{ of Ellsworth for Ellsworth in } w \text{ and} \\
& \left[ \text{win}(\gamma d) \right] \supseteq \text{Bel}(\text{ellsworth}, w)
\end{align*} \]

Truth in B is then verified with person on TV called ‘Arata’ as d. As for alternative de re LFs, a slightly more sophisticated analysis would involve a structured complement of BEL, which is then interpreted along the lines of Cresswell & von Stechow’s (1982) relational, structured propositions analysis (cf. 2.3.3.1):

(65)  
\[ \begin{align*}
a. & \text{BEL}_e^\times \text{et}(e, \text{win}) \\
b. & \left[ (65a) \right]_{w} = 1 \\
\text{iff } & \text{there is an acquaintance relation } R \text{ holding between} \\
& \text{Ellsworth and Ellsworth in } w \text{ and Ellsworth believes that} \\
& \text{the person she bears } R \text{ to is winning}
\end{align*} \]

A third possibility would be to have a nicely compositional LF interpreted according to Kaplan’s (1989) quantified character theory (cf. 2.3.4). But, for simplicity, let’s just follow the unspecified wide scope LF (63a).

The interesting part of Chierchia’s proposal is its explication of the relation between these LFs and their surface realizations. More precisely, he’s interested in where, syntactically speaking, the property-type complement in de se readings comes from. His answer starts by observing that the crucial

\[^{36}\equiv (\lambda x[\text{BEL}_x \text{win}(x)])(e)\]
difference between the *de re* and pure *de se* ifs is the extra $\lambda$ in the latter that binds the subject of the complement clause. In other words, we can get from (63a) to (62a) by inserting a $\lambda x$ to bind the deepest variable:  

$$\exists x [x = e \land \text{BEL}_x \text{win}(x)]$$  

$$\leadsto \exists x [x = e \land \text{BEL}_x^* \lambda x [\text{win}(x)]] \equiv \text{BEL}_x^* \lambda x [\text{win}(x)]$$

What happens here is we insert a $\lambda$ to shift the sentential complement into a property and since, moreover, the $\lambda$ binds the embedded variable, we get the pure *de se* reading. Note that if we had chosen a different variable, $\lambda y$, we’d have gotten a vacuous property reading, self-ascribing $[\lambda y [\text{win}(x)]]$ means self-ascribing the property of inhabiting a world where $x$ (Ellsworth) wins, which corresponds to believing the proposition that she, under whatever guise, wins. Chierchia does not consider this possibility for unification, which will be an important ingredient in later developments to be discussed in 2.4.3.

On a semantic level the difference between pure *de se* and general *de re* is one measly $\lambda x$. Unfortunately, neither coreferential pronoun reports, nor infinitivals show something that could count as the surface realization of this binder, so a hidden operator (“adjoined to CP”, in the GB implementation (Chierchia 1989:10)), let’s call it $\Lambda$, is posited. Chierchia’s GB approach thus assigns two syntactic analyses for a coreferential pronoun report:

$$\exists x [x = e \land \text{BEL}_x \text{win}(x)]$$  

$$\leadsto \exists x [x = e \land \text{BEL}_x^* \lambda x [\text{win}(x)]] \equiv \text{BEL}_x^* \lambda x [\text{win}(x)]$$

A genuine ambiguity plus a hidden operator to derive the readings we could have gotten for free on our *de re/de se* unificatory attempts of 2.3? So far, not so good, but the payoff lies in the analysis of unambiguously *de se* reports—the bane of *de re* unification—and anaphorized/elided versions thereof, which form a new set of data to be introduced in 2.4.1.4 below.

What is Chierchia’s famed analysis of infinitival and gerundial, pure *de se* reports? As we have seen before, in GB, infinitival complements are assumed to have an unrealized subject, PRO. The stipulation that is needed to enforce the right reading of infinitival reports, is that PRO *has* to be bound

---

37 Technical remark: We must set up the syntax and semantics of our formal language in such a way that a variable is bound by the closest lambda or quantifier that can bind it. This is to make sense of formulas that have two x-binders with overlapping scope.
locally, i.e. by our property abstractor $\Lambda_i$ (i.e. a $\lambda$, semantically speaking). Assuming this as the crucial difference between PRO and normal, third person, coreferential pronouns, for which this binding is only optional, gives us a correct analysis. In short:

(68) Ellsworth believes to be on the winning side.  
- Ellsworth$_i$ believes $[\Lambda_i [\text{PRO}_i \text{to be on the winning side}]]$  
  $\triangleright$ $\text{BEL}^*_x \lambda x [\text{win}(x)]$ 

In the next subsection we’ll look at some more data bringing to light interesting divergences between Chierchia’s ambiguity thesis, total unification, and intermediate proposals.

2.4.1.4 Abstract anaphora: Chierchia vs. the hybrids

The obvious criticism against Chierchia’s theory above concerns not its treatment of infinitives, but the postulation of a syntactic ambiguity in pronoun reports. Is there perhaps a way to avoid that ambiguity, while more or less maintaining the idea of PRO denoting a locally bound variable to account for infinitival pure $de$ $se$? Well, we can define truth conditions for the $de$ $re$ LFs in such a way that they cover the whole range of $de$ $re$/ $de$ $se$, in fact that’s exactly what we’ve been doing quite extensively in 2.3. If we take on Cresswell & von Stechow’s (1982) relational explication of the $de$ $re$ reading, as suggested in (65a), we get that pure $de$ $se$ asymmetrically implies $de$ $re$, i.e. pure $de$ $se$ becomes a special case of $de$ $re$, or rather, $de$ $re$/ $de$ $se$, as we have been calling it. Thus, we don’t need a separate $de$ $se$ LF or If for pronoun reports anymore, just:

(69) a. Ellsworth believes she’ll win  
  $\triangleright$ $\text{BEL}^{ext}_e (e, \text{win})$  
  $\equiv \exists R [R(e,e) \land \text{BEL}^*_x \lambda x [\text{win}(\forall y[R(x,y)])]]$  

b. Ellsworth believes to be on the winning side.  
  $\triangleright$ $\text{BEL}^*_x \lambda x [\text{win}(x)]$ 

In other words, take the old unified analysis of $de$ $re$/ $de$ $se$ for pronoun reports, and in addition assume that Chierchia’s story applies to the class of infinitival and gerundial complements with PRO subjects. This kind of hybrid account has been suggested by Reinhart (1990), although her version relies on the Boër & Lycan (1980)-style propositional unification of $de$ $re$/ $de$ $se$ that we argued against in 1.2.4.1. Another implementation of the same idea (unified account for pronoun reports but separate If for PRO reports) is developed by Zimmermann (1991) who uses Kaplan’s quantified character
Chapter 2. Attitude reports

Chierchia, however, comes up with some counterevidence, arguing against such an intermediate position and in favor of a real ambiguity in pronoun reports. These involve anaphorized belief complements:

(70) Ellsworth believes she’ll win, and her father believes that too

Intuitively, the second clause is compatible with Ellsworth’s father believing that his daughter will win, and with believing that he himself will win (the preferred reading depends, of course, on the context and the choice of markers like but instead of and). Such readings are often referred to as strict and sloppy, respectively. Strict/sloppy ambiguities are a common and well-studied phenomenon, for instance in VP ellipsis: Ligia loves her cat, and so does Essie can mean either that Essie loves Ligia’s cat (strict), or her own (sloppy). In our example, the ambiguity seems to arise from having a so-called propositional anaphor (that) refer back to (the denotation of) a phrase containing a pronoun (she).\(^{38}\) In the course of this subsection I will offer a variety of abstract anaphoric belief reports, some of which can be used as arguments in favor of Chierchian ambiguity, some against it. My final conclusion will be that the abstract anaphora data do not tip the scales either way, so apparently an additional mechanism is needed to account for the strict/sloppy ambiguities that we will encounter. I will not provide a detailed account here, but eventually we’ll get back to the data as a test case for the account to be developed in chapter 3.

Chierchia’s theory offers the following analysis of the data connected with (70). The first conjunct is ambiguous between a propositional (de re) and a property (pure de se) complement. Given some standard assumptions on abstract anaphora, the second conjunct copies (or anaphorically links) a

\(^{38}\)In addition to (70)’s pronominalization with that, we could have chosen it or this without much effect on available interpretations. In fact, anaphorization of the second belief complement could also have been achieved by means of a definite description, (ia), or a relative clause, (ib). Another way to achieve arguably the same thing is by ellipsis of the whole VP, (ic), and by constructions in between anaphora and ellipsis like (id). We classify all of these as instances of abstract anaphora, following Asher (1993):

(i) a. ...and her father believes the same thing
   b. ...and her father believes what Ellsworth believes
   c. ...and so does her father
   d. ...and her father thinks so too

We will not go into the syntactic and semantic peculiarities of each of these types of abstract anaphora, but note merely that they all exhibit the same strict/sloppy ambiguity (though, some with more bias one way or the other).
complement for the father’s belief from the semantic representation of the belief ascribed to Ellsworth in the first conjunct. If the first conjunct is read *de re*, i.e. \( \text{LF} = (67a) \), then the father’s belief complement can only pick up a singular proposition (about Ellsworth), giving the strict reading:

\[
\text{(71)} \quad [\text{Ellsworth}_i \text{ believes } \{\text{she}_i \text{ will win}\}] \text{ and } [[\text{her}_i \text{ father}]] \text{ believes that}_? \text{ too] }
\]

\[\leadsto \exists x \exists y [x = e \land y = \text{father}_o(e) \land \text{BEL}_x \text{win}(x) \land \text{BEL}_y ?] \]

If, on the other hand, Ellsworth’s belief is read purely *de se* (i.e. with intervening abstractor \( \Lambda_i \), (67b)) then the open slot in her father’s belief will pick up a property, viz. the property of winning, yielding a sloppy reading:

\[
\text{(72)} \quad [\text{Ellsworth}_i \text{ believes } \{\Lambda_i \text{ [she}_i \text{ will win]}\}] \text{ and } [[\text{her}_i \text{ father}]] \text{ believes that}_? \text{ too] }
\]

\[\leadsto \exists y [y = \text{father}_o(e) \land \text{BEL}_x^* \lambda x[\text{win}(x)] \land \text{BEL}_y ?] \]

At first sight, it seems that Chierchia’s original ambiguity provides a natural account of the strict/sloppy ambiguity of abstract anaphora reports. We will shortly come back to these predictions.

But first let’s see how a reductionist (or hybrid) account fares with these data. Let’s assume the relational account of pronoun reports, as briefly reconsidered at the start of this subsection and in depth in 2.3.1. The pre-resolution LF of (70) is:

\[
\text{(73)} \quad \exists y [y = \text{father}_o(e) \land \text{BEL}_e^{\times \text{et}}(e, \text{win}) \land \text{BEL}_y ?] 
\]

If indeed anaphora resolution takes place at this level of logical form, the only possible resolution for the anaphor (?) seems to be the structured proposition \( \langle e, \text{win} \rangle \), which would give us a strict reading:

\[\leadsto \exists y [y = \text{father}_o(e) \land \text{BEL}_e^{\times \text{et}}(e, \text{win}) \land \text{BEL}_y^{\times \text{et}}(e, \text{win})] \]
Chapter 2. Attitude reports

In other words, there’s no way to get a sloppy reading, one where the father’s belief is about himself.\(^{39,40}\) Another tentative way out would be to say that abstract anaphora is just a copying procedure in syntax, i.e. the \textit{that} goes proxy for the (suitably neutralized) phrase \textit{that she will win}. I will not pursue this suggestion any further but preliminarily conclude that Chierchia can account for the strict/sloppy data where the hybrid theory falters.

However, if we now take a closer look at the predictions of the ambiguity thesis we will find some faults with it as well. Let’s re-examine the analysis of (70) in (71) and (72); the first ascribes to Ellsworth a \textit{de re} belief about herself and her father one \textit{de re} about Ellsworth, the second ascribes them both pure \textit{de se} beliefs. And these are the only two possible readings. Though they do correspond to a strict and a sloppy reading, they don’t seem to exhaust the possible combinations of the \((\textit{de re/pure de se} \times \textit{about self/other})^{\text{Ellsworth,father}}\) grid.\(^{41}\) How about the sloppy readings where Ellsworth’s belief is not pure, but still \textit{de re} about herself while her father’s belief is either pure or impure about himself? Or those where Ellsworth’s is pure \textit{de se}, but her father’s is mere \textit{de re} about her (strict), or about himself (sloppy)? To discredit Chierchia’s solution all we need to do is make up some scenarios corresponding to these 4 extra-Chierchian readings and see if our (70) can felicitously report them. In fact, for the first couple just mentioned (sloppy, first belief mere \textit{de re}, second \textit{de re}/\textit{de se} about self) this has already been done, by Reinhart (1990) who offers the following critique of Chierchia’s ambiguity thesis:

\begin{equation}
\text{(74) Lucie}_1 \text{ thought that she}_1 \text{ sounded too aggressive and Lili thought the same thing (/believed it too).}
\end{equation}

\(^{39}\)Not even if we somehow first expand the definition of \textit{bel}^{\text{ex,ct}} to bring out the acquaintance relations, i.e. if the pre-resolution \text{I}f had been something like:

\begin{equation}
\exists z \exists R[z = \text{father of}(e) \land R(e,e) \land \text{bel}_z^x \lambda x[\text{win}(\text{ry}[R(x,y)])] \land ? \land \text{bel}_z^y]\
\end{equation}

The first \textit{?} is a half-hearted attempt to get in the fact that we should somehow look for an acquaintance relation for the father, otherwise we’d end up ascribing Ellsworth’s father the belief that the person he’s \textit{R}-acquainted with will win, where \textit{R} is the acquaintance relation associated with Ellsworth and her belief. I’m not sure there’s any feasible derivation of the sloppy reading along these lines. Also, even if we can get it, now the strict one has become even more problematic.

\(^{40}\)A more sophisticated semantic analysis of abstract anaphora based on Dalrymple \textit{et al.}’s (1991) account of VP ellipsis, using \textit{higher order unification} to derive strict/sloppy ambiguities, might be of help. I will not pursue this suggestion any further, but we’ll get back to the higher order unification tool in later chapters.

\(^{41}\)Well, the parameters are not really independent, e.g. pure \textit{de se} is always about self, discarding 7 of the total 16 already. We can get rid of 3 more by assuming, as we have been doing all along, that the pronoun in Ellsworth’s belief refers back to Ellsworth.
Lili thought she sounded too aggressive (ambiguous).

Suppose that both Lucie and Lili were recorded, and, not recognizing their voice, they each thought that it sounded too aggressive. It is perfectly possible to report this state of affairs using (74), in which case Lili’s belief is construed as in (75). But this is a clear case of a de-re belief. If the construal in (75) was allowed only in the case of de-se belief as argued by Chierchia, this reading could not be obtained. In other words, the construal (75) [...] is still ambiguous between the de-re and de-se interpretation (Reinhart 1990:4)

I will leave it to the reader to find out if the remaining (two) readings can also occur.

It seems that neither a unificatory hybrid, nor Chierchian ambiguity can by itself derive all available strict and sloppy readings of (70), which leads me to conclude, with Reinhart, “that the anaphoric process illustrated in these contexts is independent of the de-se issue.” (Reinhart 1990:4). I would suggest a combination of the relational hybrid with a higher-order unification analysis of abstract anaphora and ellipsis resolution.

Let’s move on to Chierchia’s next pro-ambiguity argument based on anaphorized belief complements. According to his intuitions:

the sloppy reading disappears if the antecedent [of the abstract anaphor] is taken to be a pronounless structure like the one in (76):

(76) a. Domingo_i believes that Domingo_i is a genius
b. Domingo_i believes that [the author of this book]_i is a genius

Sentences of this kind are slightly odd due to a principle C effect [...] However it has been noted that these sentences can be

Reinhart fleshes out her recorded voice version of the worn pants-on-fire mistaken identity example as follows:

(i) Lucie thought that she sounded too aggressive.

[...] Suppose that Lucie, a broadcast manager, is looking for the perfect female voice for an ad, and requests to hear some samples of women in natural conversation. Unbeknown to her, the technician records her too, and adds it to the samples as number 17. Lucie does not recognize her recorded voice, and rules out 17 as too aggressive. In this scenario, (i) is true de-re, but is still false de-se (for the specific reported thought). (Reinhart 1990:1)
rescued in suitably contrastive contexts [...]. Yet, even when they are grammatical, such sentences never support sloppy anaphora [...]

The “principle C violation” can be made more palatable, in the case of (76a), by setting up the context so as to put extra focus on either of the two occurrences (*everybody thinks Domingo’s a genius, even Domingo thinks Domingo is a genius*), or, for (76b) (and in general in cases of different but coreferring referential expressions), by placing it in a mistaken self-identity context. So I agree with Chierchia that these sentences are grammatical, but he goes further and considers the data evidence for his ambiguity thesis, which, he says, excludes a (pure) *de se* interpretation of such sentences, presumably because his property-abstracter can only bind pronouns, the surface realizations of variables, not ‘R(eferring)-expressions’ like definite descriptions or proper names.

But how robust are these intuitions? And what would a unificatory analysis predict? Let’s focus on the repeated proper name example first:

(77) [Very few people believe that Domingo is a genius. But,] Domingo believes that Domingo is a genius.

\[
\begin{align*}
\text{Pavarotti believes that as well} \\
\text{so does Pavarotti}
\end{align*}
\]

But then,

\[
\begin{align*}
\text{his rival believes the same thing} \\
\text{everybody thinks that, sometimes} \\
\text{don’t we all?}
\end{align*}
\]

I’m not at all convinced such discourses require strict readings of the second part. If indeed some of the above variations are true if Pavarotti or ‘everybody’ merely thinks “I am a genius”, we’d have another counterexample to Chierchia’s theory. As for the relational analysis, well, pending an independent analysis of sloppy anaphora, we do already predict that, regardless of Principle C violations, the antecedent belief report is true in *de re* as well as pure *de se* scenarios (contra Chierchia). It is thus to be expected that the relational analysis will allow sloppy readings. Note that if this prediction is not borne out, we still have the option of blaming the Principle C effect itself for blocking the sloppy anaphora. I envisage a (pragmatic) argument along the following lines: because the proper name repetition is quite unusual/marked, it makes the name and its bearer so salient that the consequent ellipsis is automatically taken to still be about that same individual, thus forcing a strict reading.

---

43Bart Geurts (p.c.) and Jennifer Spenader (p.c.) note that principle C violations like (76a) also occur quite naturally in legalese and motherese respectively.
2.4 Anti-reductionism and pure de se separatism

Elaborating on this issue, let’s extend the range of examples a bit. Applying Chierchia’s reasoning to our earlier (1b) (Ellsworth believes of Arata Suggs that she’ll win) we’d get that it does not have a pure de se reading, i.e., following Chierchia’s view of de re, it’s false in scene C where Ellsworth/Arata sincerely says: “I am going to win.” As we saw in table 2.4 in 2.3.1, the relational account makes it true there, because whenever the matrix and embedded subjects corefer we get de re/de se. I am not sure which prediction is correct, but it does seem correct that in the progression below the pure de se reading becomes less and less prominent, in favor of (mistaken-identity) de re or de dicto readings (cf. table 2.1 p. 75, where we already put question marks for these last two judgments):

(78) Ellsworth = Arata (though she doesn’t know that nickname) = the prettiest (though she’d disagree), says: “I am gonna win” [≈C, p. 74]

Ellsworth believes
a. PRO to be winning [≈(59b)]
b. she’ll win [≈(59a)]
c. Ellsworth will win [cf. (76a)]
d. ?Arata will win [≈(1b), p. 72]
e. ??the prettiest will win [≈(1b), p. 72]

I have yet to see a fully worked-out account of this, but I strongly disagree with Chierchia who draws a sharp line between (78b) and (78c). In defense of the rather weak relational predictions I can only suggest that we can always augment our semantic analysis (all are true, semantically speaking) with a pragmatic explanation of the infelicity of the lower ones. Pragmatic explanations could include the fact that the reporter’s word choice in (78d-e) is quite misleading; if Ellsworth talks about herself in the first person why would a report use two different names? A somewhat different, yet related, pragmatic explanation could build on the uncontroversial assumption that the use of proper names and descriptions is blocked by ‘lighter’ semantically compatible alternatives (in this case PRO and the pronoun in (78a-b), which can already express a pure de se). This in turn already comes pretty close to the above pragmatic reasoning using the markedness to block the sloppy anaphora.

---

44 Actually, Chierchia is not too clear and seems to change his mind about whether or not the de re LF corresponds to the relational analysis’ unified de re/de se, or whether it’s dedicated to de re plus impure de se. I put Chierchia in the latter camp, but note that the former view is suggested by his discussion of a meaning postulate that turns de se into a subset of de re.
Getting back to the possibility of sloppy anaphora, which on Chierchia’s account depends on the availability of a pure de se LF, i.e. only (78a-b) support sloppy readings of anaphorized continuations. I agree that sloppy readings for (78d-e) are out, but for (78c), see (77) above, and for (78a), see (79) below. Before moving on to this anaphoric continuation of (78a), let me put all this criticism in perspective by reminding you once again that the relational (hybrid) alternative by itself still has no way to account for any sloppy readings whatsoever, so we’re arguably hardly better off than with Chierchia. The fact that Chierchia gets the facts wrong as well, however, I take as purely negative evidence that more than just de re/de se is needed to account for these (sloppiness) data. As suggested before, a separate ellipsis/abstract anaphora mechanism is called for.

The last piece of data to be discussed here concerns an unexpected strict reading in the dual of the above proper name example: a report that is unambiguously pure de se serving as antecedent for an abstract anaphor:

(79) Ellsworth believes to be on the winning side, but her father doesn’t believe that. He thinks she should have stuck with the other team.

This discourse seems fine, so the that must have picked up a de re belief, about Ellsworth. Chierchia fails to account for this strict reading, this time because he only assigns a property to the PRO-headed complement of the first belief. Again, the problem remains for the hybrid analysis, precisely because it coincides with Chierchia’s on PRO.

This concludes the discussion on unambiguously pure de se reports. Although we’ve had to deviate from our original intention to focus solely on the attitude of belief in English, the data, featuring indirect reflexives (2.4.1.1), logophors (2.4.1.2) and PRO+infinitive (2.4.1.3), do constitute a genuine

45This judgment of a strict/sloppy ambiguity with infinitival pure de se antecedents is attested by quite a few (mostly Dutch) people, e.g. in response to this homework assignment (B3, Fall 2004):

(i) Marie wil gelukkig zijn en haar man wil dat
  Mary wants PRO happy be.INF and her husband wants that
  ook
  too
  ‘Mary wants to be happy and her husband wants that too’

How many readings do you get (intuitively)? Can you get a “strict” reading for the elided content of the husband’s wants? Does anything change for your intuitions if you replace “that too” with “it too”, “the same thing”, “what his wife wants” or something similar?

46Mysteriously, Chierchia himself recognizes that this strict reading exists (p.23), but somehow he doesn’t consider it to be problematic.
2.4 Anti-reductionism and pure de se separatism

problem for de re/de se unification. One way to take care of it is to abandon the unity of de re/de se and follow Chierchia’s ambiguity thesis, according to which pronouns in belief complements make reports syntactically ambiguous (2.4.1.3). In 2.4.1.4, hybrid theories, taking an intermediate position between unification and ambiguity, were shown to be a viable alternative, despite prima facie counterevidence based on strict and sloppy readings of abstract anaphoric reports. In 2.4.3 we will introduce more data (shifted indexicals) that are supposed to put an end to both Chierchia’s and hybrid unificatory analyses (together classified under ‘intermediate’ theories by Schlenker (2003)). My own analysis in chapter 3 offers a significant improvement over the simple hybrid sketched here in that it provides a unified account of de re/de se reports of the form NP believes that NP VP that treats embedded PRO truly on a par with anaphoric and (shiftable and unshiftable) deictic expressions (though some stipulations about the special semantics of PRO are unavoidable, cf. 2.4.3.3 and 3.4.4.4).

2.4.2 De re and de se under quantifiers

In this section we discuss more data in connection with the de re/de se ambiguity vs. unification debate. This time it’s quantified belief reports, i.e. sentences like everybody believes they will be chosen and nobody thought that Arata Sugs would win. Though there is quite compelling evidence to the contrary (non-conservative quantifiers anyone?), we will include here only as a quantifier, following e.g. de Mey (1991), but mainly in order to reconstruct a particular argument against de re/de se unification depending on it (in 2.4.2.2).

As in the previous section, we will end up with judgments that can be accounted for in simple reductionist frameworks but not in the Chierchian LF ambiguity framework, and vice versa. We will discuss the effects of some small framework tweaks, but you’ll have to wait until the next chapter, which brings first a more serious and modern treatment of context dependence, for a framework that can handle them all.

2.4.2.1 Zimmermann: every

In this chapter we have been looking at reports of the form NP$_1$ believes that NP$_2$ VP with NP$_1$ a proper name and NP$_2$ a proper name, description, indexical, anaphor, logophor, or a PRO. It turned out that the most interesting data were these last three where both NPs are coreferential, because that’s where the de re/ vs. pure de se distinction comes to the fore. Now that we’ve gotten bored tampering with the res denoting second NP, it’s time to play
around with NP₁. In this subsection we consider a universally quantified variant of our much discussed (3), *Ellsworth believes she’s going to win*, in scenarios with multiple believers.

The first discussion of such data is by Zimmermann (p.c.) whose example (allegedly)⁴⁷ involves several drunk election candidates, some of whom saying, “I should be elected”, and others saying “She should be elected” about themselves. According to Zimmermann, spectators might felicitously describe this situation by saying *Each candidate hopes that he gets elected*.

For simplicity’s and uniformity’s sake, let’s strip it down to the bare essentials (two individuals, *believe* instead of *hope*) and reformulate it in terms of the ongoing Ellsworth Kimmel story of 2.1:

D. Ellsworth Kimmel and her rival Shavonne McManus are at the party described earlier. Not realizing *Arata Suggs* is used as a nickname referring to herself, Ellsworth tells Shavonne: “That Arata Suggs person on TV is probably going to win”, to which Shavonne responds with a “Fat chance! I’m gonna win.”

A minimal scenario, in a sense combining scene B, where Ellsworth had an impure de re/de se belief about herself, with scene C, where the belief (here, Shavonne’s) is impure, but still about herself. From 2.1 we know that belief reports with coreferential third person pronouns can report both of these types of beliefs, so (80a) and (80b) are both true (even if the second is much harder to read as true), so it seems natural to combine them into a single report, (80c):

(80)  a. Shavonne thinks she’s going to win  [≈(3), p. 73]
   b. Ellsworth thinks she’s going to win  [= (3), p. 73]
   c. Both women think they’re going to win

Indeed the quantified version seems as true as a mere conjunction of (80a) and (80b).

Note that we’ve already discussed an alternative way to conjoin coreferential pronoun reports, viz. by ellipsis/abstract anaphora of the second, which led to some interesting observations in 2.4.1.4:

(81)  a. Shavonne thinks she’s going to win and so does Ellsworth
   b. Ellsworth thinks she’s going to win and Shavonne believes that

⁴⁷Schlenker attributes it to his 1991 and others have followed this attribution. Reinhart (1990) seems to have a similar intuition, though she does not really give a mixed scenario. Nor does Chierchia (1989), but the truth conditions he suggests would make the sentence false in a mixed context, contradicting Zimmermann.
As described in 2.4.1.4, the first to consider such conjoined reports was Chierchia (1989). We’ve already criticized his account for predicting that a pure \textit{de se} antecedent implies a pure (sloppy) reading for the elided conjunct (so (81a) becomes false in D), and for predicting only strict readings when the antecedent is just \textit{de re} (so (81b) is also false). In passing, Chierchia also mentions quantified reports like (80c):

The very same contrast [between \textit{de re} and pure \textit{de se} LFs] obtains, of course, if we have quantified NPs. Thus consider, for example, (82).

(82) Everyone in that room thinks that he is Hume

... 

The most plausible interpretation of (82) claims that each person in the relevant room has a certain [pure] \textit{de se} attitude (perhaps due to schizophrenia). (Chierchia 1989:10)

Ergo, (80c) is false in D.\(^{48}\)

In a reaction to that paper, especially to his judgments concerning reports of the form of (81b), Reinhart (1990) claims that sloppy readings are possible whether or not the antecedent of the abstract anaphor is read purely \textit{de se}, and moreover, if the antecedent is not pure, the second conjunct’s sloppy reading is ambiguous between pure and general \textit{de re}/\textit{de se}, so (81b) would come out true. It’s unclear what she’d say about (81a) in D, but the LF assigned to (80c) in analogy with (81b) is true.

So much for an informal discussion of the data, let’s test our frameworks on it. Chierchia’s ambiguity obviously fails here. For the anaphorized reports this failure has been documented in detail in 2.4.1.4 above. For the universal quantification example we predict that either there is an appropriate \(\Lambda_i\) binder (and then the truth conditions are such that both believe “I am going to win”), or else, there’s no intervening \(\Lambda_i\) binding the embedded pronoun and both beliefs are \textit{de re} (and as before, I take it that that means propositional/non-\textit{de se}, cf. 2.4.1.3), which is equally false in D:

(83) Both women believe they are going to win

\(\approx(80c)\)

\[
\begin{align*}
\text{a. } & \left[\text{[both women]}, \ [\text{believe} [\Lambda_i [\text{[they}, \ i \text{are going to win]]}] \right]
\text{ } \\
& \rightarrow \forall x [\text{woman}(x) \rightarrow \text{BEL}^*_x \lambda y [\text{win}(y)]]
\end{align*}
\]

\(^{48}\)Note that I do not really consider this a semantic intuition about the truth value of a sentence in a specific context, rather I’d say it’s a faulty prediction of the theory.
b. [both women]i [believe [theyi are going to win]]
\[\forall x[\text{woman}(x) \rightarrow \text{BEL}_z\text{win}(\tau y[R(x, y)])]]\]

Simple unification of de re/de se (relational or characterial, hybrid or not) on the other hand works out of the box for the quantified reports like (80c):

\[(84) \quad \forall x[\text{woman}(x) \rightarrow \text{BEL}_z(x, \text{win})] \equiv \forall x[\text{woman}(x) \rightarrow \exists R[R(x, x) \land \text{BEL}_z\lambda y[\text{win}(\tau z[R(y, z)])]]]\]

In words, for each of the women there should be some mode of presentation of the res (= that woman herself) under which the belief holds. It’s easy to see that this works and, by the same token, so would a (fully compositional) characterial analysis. Recall however that we had some trouble with the abstract anaphoric examples in 2.4.1.4 and the conclusion there was that the relational theory as such can’t handle sloppy anaphora. On the other hand, since sloppy anaphora occur outside the realm of belief reporting anyway, this was hardly surprising and it seems likely that an independent account of sloppy anaphora (say, along the lines of Dalrymple et al.’s (1991) higher-order unification account of ellipsis) will do the trick.

Contextualizing the acquaintance relation along the lines of Abusch (1997), among others, seemed like a good idea in 2.3.2, but here we hit upon a serious drawback. What salient acquaintance relation do we pick from a mixed context? Obviously, picking either one gives the wrong result: either we take up Shavonne’s acquaintance relation and we end up saying that both are consciously acquainted with themselves as themselves (pure de se), or we pick Ellsworth’s so that both have to see themselves on TV and believe that the person they are currently seeing will win.\(^{49}\) If we think in terms of scope, we must conclude that in mixed contexts the acquaintance relation cannot outscope overt subject quantifiers like the one contributed by both of (80c). This is a significant hurdle in the way of a unified, pragmatic account of de re/de se, which the current work attempts to provide. It will be overcome in 3.4.3.

To sum up this section, here’s a table. The first two rows are old news, repeated for completeness. The data with respect to the next two, the elliptic constructions, are extrapolated from remarks by Reinhart (1990) and my own discussion in 2.4.1.4. In the last row, the quantified report, the data is Zimmermann’s rather uncontroversial judgment. I take it then that

\(^{49}\)Even worse, if Ellsworth’s acquaintance to the res involves the fact that people are referring to it as ‘Arata Suggs’, the resulting readings for (80c)-(81b) under the contextual acquaintance theory will end up ‘strict’, saying in a sense that both have beliefs about the same res, implying moreover that both women are actually referred to by that name.
Chierchia’s remark about an analogous quantified report (quoted above) is more a (faulty) prediction of his theory than a genuine, robust intuition that extends to mixed utterance contexts like this one.

<table>
<thead>
<tr>
<th>D. data</th>
<th>Chierchian ambiguity</th>
<th>de re/de se unification</th>
<th>contextualized acquaintance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(80a)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(80b)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(80c)</td>
<td>1</td>
<td>0</td>
<td>1?</td>
</tr>
<tr>
<td>(81a)</td>
<td>1</td>
<td>0</td>
<td>1?</td>
</tr>
<tr>
<td>(81b)</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.6: Intuitions and predictions in mixed scenario D

### 2.4.2.2 De re/de se asymmetries with only

The previous subsection showed how Zimmermann’s example can be used to argue for a simple (characterial) unification of de re/de se. Ironically, the same scenario, with a different sentence, has been used as an explicit argument for pure de se separatism. The current subsection introduces, discusses and extends that argument, which is due to Percus & Sauerland (2003a).

The data they present involves a mixed scenario, so let’s use our D, in which Shavonne has a first person belief and Ellsworth a third person belief about herself. The sentence they judge true in D is:

(85) Only Shavonne believes she will win

Let’s grant that this is indeed the case and consider the consequences for a theory of de re and de se belief reports. As before, we’ll check the predictions of the main frameworks under discussion here, to wit (i) simple unification via acquaintance or characterial quantification, (ii) de se purist separatism, and (iii) contextualized unification-by-acquaintance.

However, before attempting this comparison we must make some decisions about the semantics of only, a notoriously ill-behaved adverb. It has been a matter of much debate what semantic and/or pragmatic category or categories only belongs to and it is certainly not my intention to add anything to it. Historically, only was first analyzed as, or in terms of, quantification. Its meaning was something like the reverse of the universal quantifier: only

---

50See for an overview of the “classic” positions http://staff.science.uva.nl/~pdekker/Only/
philosophers understand this ≈ everybody who understands this is a philosopher. In generalized quantifier theory quantifiers are construed as relations between two sets, called restrictor and (nuclear) scope, as in some\((A,B)\), where \(A\) is the restrictor (or domain of quantification) and \(B\) the scope, meaning “some A’s are B” and definable as \(A \cap B \neq \emptyset\). The relation between all and only comes out as \(only(A,B) = all(B,A)\). Sure enough the literature contains many objections and complications connected with the analysis of only as a reversed universal quantifier, but none of these will concern us here, just to keep things manageable. One way of defending this simplification would be to say that we’re currently interested in representing truth conditions, not in, say, determining how these truth conditions depend on focus patterns, or what aspects of only’s meaning are truly semantic, (conventionally or conversationally) implicated, and/or presuppositional. To sum up then, I propose the simplest analysis of the quantifier only imaginable, and represent it as follows:

\[
\begin{align*}
(86) \quad \text{Only NP}_P \text{ VP}_Q \\
\quad \triangleright \quad \text{ONLY}_x [P(x)] [Q(x)] \\
\quad \equiv \quad \forall x [Q(x) \rightarrow P(x)]
\end{align*}
\]

Now, we’ll be focusing on sentences where the restrictor set \(P\) is given by a definite NP, which are most naturally represented as in (87) (which employs the relational account in the analysis of the quantifier’s scope):

\[
\begin{align*}
(87) \quad \text{Only Shavonne believes she will win} \\
\quad \triangleright \quad \text{ONLY}_x [x = s] [\text{BEL}_x (x, \text{win})] \\
\quad \equiv \quad \forall x [\exists z [R(x, z) \land \text{BEL}_x^* x y [\text{win}(\text{ι} z [R(y, z)])] \rightarrow x = s]]
\end{align*}
\]

As with every in 2.4.2.1, this straightforward implementation of de re/de se unification gives a reading where the existentially quantified description of the res takes narrow scope with respect to the overt quantifier only. Note again that exactly the same would hold of the Kaplanian unification with characters as modes of presentation. We might paraphrase the unificationist truth conditions as Shavonne is the only person with the property of being acquainted with oneself in such a way that believing the person one is so acquainted with will win. In D, this is false, since Ellsworth has a belief about herself under an acquaintance relation too, except it’s not equality, but seeing a fellow contestant on TV called ‘Arata’. Thus, simple reductionist frameworks cannot account for the observation that the sentence is in fact true in D.

As Percus & Sauerland (2003a) point out, the facts are easily captured
2.4 Anti-reductionism and pure de se separatism

if we adopt Chierchia’s ambiguity thesis. Their version has it that a VP like *thinks that she will win* “has an LF with the ‘[pure] de se denotation’ [...] in addition to any LF it might have in which the pronoun is construed de re” (p.5).\footnote{51Furthermore, Percus & Sauerland (2003a) assume that de re LF is compatible with the whole de re/de se spectrum, something that was not quite clear with Chierchia (1989) (cf. discussion about this in 2.4.1.3, and, especially, footnote 44).} According to this separatist proposal the sentence may well have a de re LF, interpreted along the lines of (87) (which would be false), but in addition it will also have a distinct pure de se LF, interpreted as in (88).

\[(88) \quad \text{ONLY}_x[x = s][\text{BEL}_s^* \lambda y[\text{win}(y)]] \]

This says that Shavonne is the only person who self-ascribes the property of winning, which is true in D, because Ellsworth merely believes that the person she is seeing on TV will win. So, postulating a syntactic ambiguity in the belief reporting construction between a strong pure de se (though under only that ‘strong’ becomes weak and vice versa) and a general de re, can account for the fact that there’s a true reading in scenario D. Add to this the need for separate de se LFS for PRO constructions, plus (some of) Chierchia’s arguments with abstract anaphora and it becomes clear why the seemingly superfluous postulation of syntactic ambiguity for reports with pronouns is not that implausible.

A third possibility that needs to be considered, especially since it made such a big difference for the universal quantifier in the previous subsection, is the Hintikka-Aloni-Abusch inspired contextualization of unification by acquaintance. This is the idea that the acquaintance relation of the reductionist proposal should be determined in the global context rather than being quantified away. Not being all that interested in the semantics-pragmatics distinction at this point it makes little difference whether we implement this by leaving a free variable for \(R\) or by existentially quantifying over it with widest scope. The logical form would have to be (89), though the exact derivation is not unproblematic, depending on some details in the analysis of only \((R(x,x) \text{ vs } R(s,s), \text{ cf. 3.4.3})\).

\[(89) \quad \exists R[R(s,s) \land \ldots \land \text{ONLY}_x[x = s][\text{BEL}_s^* \lambda y[\text{win}(y)]]] \equiv \exists R[R(s,s) \land \ldots \land \forall x[\text{BEL}_s^* \lambda y[\text{win}(y)]] \rightarrow x = s] \]

Meaning, in the context Shavonne must be known to bear a certain salient acquaintance relation to herself, and she is claimed to be the only person believing to win under that acquaintance relation. Unlike the quantified, narrow scope unification considered first, this gives the correct prediction, because the context provides us with equality as a suitable acquaintance
relation relating Shavonne to herself and verifying that she’s the only one who believes to be winning under that pure *de se* acquaintance relation. Contextualization seems to be the way out of the Percus & Sauerland (2003a) argument for *de se* LF separatism.\(^{52}\)

Unfortunately, Henk Zeevat (p.c.) has pointed out that the wide scope (or contextual) acquaintance analysis sketched above fails for only embedded reports in general. Consider, still in D, the following minimal variant of (85):

(90) Only Ellsworth believes she will win

With a bit of imagination, we have judged it true that Ellsworth believes (of herself) that she will win, but in the context at hand she is not the only one who believes that: intuitions are very clear that (90) is false in D. It’s easily checked that (narrow scope) quantified acquaintance gets this right, as does the ambiguity analysis (Ellsworth is neither the only one with a general *de re/de se* belief, nor (the only) one with a pure *de se* belief).\(^{53}\) But if we extend the relational analysis to look for the acquaintance relation in the context we predict that the sentence is true in D. This is because in that context there is a highly salient and suitable R relating Ellsworth to herself, viz. the relation of seeing a rival contestant, ‘Arata’, on the TV. Taking that relation as the description under which the belief is held, we end up with a true belief of Ellsworth’s.

<table>
<thead>
<tr>
<th>D.</th>
<th>data</th>
<th><em>de re/de se</em> unification</th>
<th>P&amp;S ambiguity</th>
<th>contextualized acquaintance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(85)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(90)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.7: *Only* in mixed scenario D

Looking at the bigger picture, the minimal pair consisting of Percus & Sauerland’s (2003a) and Zeevat’s (p.c.) reports in Zimmermann’s mixed scenario shows that the acquaintance relation of equality is somehow special among the acquaintance relations you can bear to yourself. From a philosophical point of view this is perhaps not so surprising, as the acquaintance relation of equality represents the perspective you have on yourself *qua* the person you are. What’s more interesting is that these examples show that

\(^{52}\)In fact, this was the line I took in my first talk on the subject (Szklarska Poreba 2004).

\(^{53}\)But the way we’ve construed Chierchia’s original ambiguity (impure vs pure *de se*) would incorrectly make the Zeevat report true.
2.4 Anti-reductionism and pure de se separatism

the epistemologically special status of the self-perspective is grammaticalized into the language of attitude reporting, in line with Chierchia’s observations about PRO-headed reports as encoding that special epistemic perspective in natural language. Once we have this only pair, we can see parallels with even reports in D:

(91)  
   a. [Shavonne thinks she’s going to win.]
      Even Ellsworth thinks she’s going to win.
   b. [Ellsworth thinks she’s going to win.]
      #Even Shavonne thinks she’s going to win.

This re-affirms the observed asymmetry between pure de se and other forms of de re that provides the main challenge for the unificationist program.

2.4.3 Shifted indexicals and monsters

In this section we take a look at some of the data and arguments that have lately received quite some attention as the final nail in Kaplan’s monster-free coffin. The data show blatant violations of the prohibition on monsters; in some languages the reference of proven indexicals (so not just tenses) can be affected by embedding under attitudes, which show that the attitudes are monstrous operators. Although Kaplanian character theory is not our unification of choice (given the recent arguments presented in 2.3.5), I show that the data can be wielded against relational analyses as well.

As for the data, we had promised to concentrate on the English language, the attitude of belief, and res in the person domain. In order to present these highly relevant and influential arguments we must broaden our interests with respect to the language or the domain parameters. As in the discussion of logophoricity in 2.4.1.2, we will include other languages, mainly because the examples involving modalities other than person, such as tense and mood, are generally less clear and, in the case of tense, readily explained away on independent grounds (e.g. anaphoricity of (English present) tense, some obnoxious sequence of tense phenomena).

Schlenker (1999) was the first to challenge Kaplan’s monster prohibition on the basis of empirical, linguistic evidence.\textsuperscript{55} He sets out to prove Kaplan wrong by showing that natural languages do contain monstrous operators, viz. the language’s attitude reporting constructions. The first thing to point out was that attitude reports are not always underspecified for the de re or

\textsuperscript{54}Brought to my attention by Michael Franke.
\textsuperscript{55}Perry & Israel’s (1994) earlier ‘plea for monsters’ was based solely on theoretical arguments.
de se mode of presentation of the reported attitude, contrary to Kaplan’s (and other) unified semantics’ predictions. Complementing the mere theoretical arguments to this effect, e.g. by Castañeda (1966), Schlenker cites actual data showing unambiguously pure de se reports. These data are, first, the African logophors, discussed in 2.4.1.2, and second, Chierchia’s PRO in English, 2.4.1.3.

He briefly considers a so-called ‘intermediate theory’ to account for the data so far. His intermediate proposal is basically a reformulation of Chierchia’s ambiguity analysis and of what we have called hybrids in 2.4.1.4. It is a theory which avoids monstrous belief operators by postulating that attitude operators can bind certain pronominal items, notably logophors and PRO, but not ‘real’ indexicals like English I, for that would contradict the Fixity Thesis (and therefore the ban on monsters, and Principle 2, cf. 2.3.4). Completely in line with Chierchia’s analysis in 2.4.1.3, PRO is represented as a locally bound variable, while I is free (or else at least not bound by/inside the attitude). There are several ways to represent this idea formally: (i) we could turn $\text{BEL}^{\text{mat}}$ itself into a quantifier, explicitly quantifying over individuals, $\text{BEL}_{e}^{\text{mat}} x \phi$, so it can bind PRO, or (ii), staying closer to Chierchia’s and others’ formulas, we could stick with the self-ascription operator $\text{BEL}^{*}$ taking property-type complements, prefixing the complement with a $\lambda$, and bind any logophoric variables in it. Continuing in this last, Chierchian fashion, we get:

\[ (92) \quad \begin{align*}
\text{Ellsworth believes} \quad \left\{ \begin{array}{l}
\text{she}^{*} \text{ is LOG be PRO to be}
\end{array} \right\} \text{ on the winning side}
\end{align*} \]

\[ \updownarrow \quad \text{BEL}_{e}^{*} \lambda x [\text{win}(x)] \]

\[ \begin{align*}
\text{Ellsworth believes} \quad \left\{ \begin{array}{l}
\text{she}
\end{array} \right\} \text{ is winning}
\end{align*} \]

\[ \updownarrow \quad \text{BEL}_{e}^{*} \lambda x [\text{win}(e)] \]

Various bells and whistles will be added below, and when we pick up this framework in 2.4.3.5.\textsuperscript{56} The essential ingredient however remains that logophoric pronouns and PRO are never treated exactly like (first person) indexicals: indexicals get their reference from the context, locally bound

\textsuperscript{56} E.g. explicitly representing (and abstracting over) person, tense and mood; unifying matrix and embedded clauses as denoting properties; unifying de se pronouns and indexicals as variables at lf.
variables do not. And, frankly, that seems quite alright, and, for reasons of conservativity preferable over any monster-based account.

So far so good for the classical theory, but at this point Schlenker brings in a new set of data showing real indexicals blatantly violating Principle 2. These *shifted indexicals* occur in various forms of indirect discourse and the paradigmatic example is Amharic’s\(^{57}\) shifted first person:

\[
\text{(93) }
\begin{array}{c}
\text{jon } \text{gna } \text{a} \text{-ni } \text{yil } \text{-all}
\end{array}
\begin{array}{c}
\text{john hero be -1.sg say.3.sg -aux.3.sg}
\end{array}
\quad \text{Amharic}
\]

\[
\begin{array}{c}
\text{‘John, says that he, is a hero’}
\end{array}
\quad \text{(Schlenker 2003:68)}
\]

The embedded first person morpheme is rendered with a third person in the English translation, since it refers to the matrix subject John, not to the actual speaker. At first sight, the data appears to be in direct violation of Kaplan’s Principle 2, that indexicals are always interpreted with respect to the utterance context. It remains to be seen whether and how this necessitates monsters, but, in any case, Amharic contrasts with English, where the first person pronoun’s reference is always determined by the speech context, independent of modal or attitudinal embeddings.

So what is going on here? There are a couple of ways to deal with these data, one being the introduction of monsters, but to maximize dramatic impact let’s save that one for last. First let’s debunk some simpler alternatives: shifted indexicals as anaphors 2.4.3.1, and shifting as quotation 2.4.3.2. After presenting Schlenker’s empirically superior yet monstrous analysis in 2.4.3.3-2.4.3.4, we will investigate whether that account really is monstrous by comparing it with von Stechow’s (2002) allegedly nonmonstrous variant in 2.4.3.5.

### 2.4.3.1 Indexicals as pseudo-indexical anaphors

The first option that springs to mind is to deny that Amharic *I* is indexical, in which case Kaplan’s prohibition of monsters would be saved and we could stick with an intermediate theory, analyzing this Amharic first person as we did logophors and PRO: as anaphors/variables bound by a lambda. In 2.4.3.5 we’ll outline a more fully worked out semantics based on the idea of Amharic *I* as a ‘fake indexical’ (Kratzer 2006), but here let’s consider the general arguments against such a workaround.

The first counterargument is of a merely theoretical nature, concerning the danger of circularity. It urges us to be careful not to take Principle 2 as

---

\(^{57}\)A Semitic language of Ethiopia with over 17 million speakers.
an a priori rule and argue that Amharic first person (and any other shiftable items found in other languages and domains, see examples below) cannot be indexical because of Principle 2. So construed Kaplan can never be falsified. With Schlenker (2003:31-2), I think we should view it rather as an empirical generalization that makes for a nice analysis of the classical, introspective data set, but, perhaps, does not stand up to closer, cross-linguistic scrutiny. So construed Kaplan can never be falsified.

A second, more empirically driven counterargument builds on the observation that the Amharic first person pronoun occurs in main clauses and non-reporting embeddings as well. This already sets it apart from African logophors and controlled PRO which, as observed in 2.4.1, are confined to (various forms of) indirect discourse. One might still argue then that Amharic I is an all-purpose anaphor that can be bound by the Chierchian lambda, but, in other types of clauses also by other, salient agents. Note however that in all but the report-embedded cases, the Amharic first person morpheme is used indistinguishably from English I:

\[(94) \text{ wəndim } -\text{ yəmm } -\text{ wəd} -\text{ at } -\text{ in } \text{lījīj aгоłnə } \\
\text{ brother } -\text{ pos1.sg rel } -\text{ 1-love -3.fem -acc girl find.3m } \\
\text{ ‘My brother found a girl I like’ [Amharic (Schlenker 2003:69)]}\]

Crucially, this sentence cannot be used to mean ‘My brother found a girl he likes’, which rules out analyses of Amharic first person as a feature-free anaphor that can pick up any salient, animate individual from the context. In fact, even when embedded in speech reports, the Amharic I of (93) still has a strict reading, as in John says that I am a hero. So, assimilating the

---

58 There’s some discussion about the theoretical status of Principle 2 and the Prohibition of Monsters by Zimmermann (1991:167-8), who tries to defend Kaplan by applying the hack we’re currently arguing against. In the opposite camp, Perry & Israel (1994) argue that Kaplan’s Prohibition is pure stipulation that should be done away with. However, they do not provide any new empirical evidence for their claim, nor a semantics that could compete with Kaplan’s on the familiar English introspective data. Then, after Schlenker’s data had become well known, Zimmermann’s original discussion is picked up by von Stechow & Zimmermann (2004:fn.11,15), who now stress they take the monster ban to be a falsifiable, empirical claim. Interestingly, like Zimmermann (1991), von Stechow (2001) still argues against Schlenker’s monstrous analysis, instead trying to save the ban on monsters, by removing semantically indexical features from Amharic I, a version of the intermediate theory to be discussed in 2.4.3.5.

59 At least, as Schlenker conceives of logophors. On p. 117 I briefly questioned the validity of the claim that logophors only occur in reports with (56).

60 I am ignoring the kind of PRO that is hidden in to win would be nice.

61 I haven’t seen actual glosses that show that (93) has this ‘Kaplanian’ reading, but it is certainly implied by Schlenker’s discussion, and predicted by his analyses. von Stechow (2002:8) makes it explicit by writing that in Amharic “an embedded first person pronoun can refer to the [actual] speaker or it can be interpreted de se.”
Amharic shifted first person to logophoricity and control would introduce unwanted ambiguities: Amharic I would be the equivalent of an English I in some cases, or something completely different in others. A unified analysis as free variable/anaphor overgenerates: there would be no way to account for the fact that Amharic I is shiftable only in reports. It is perhaps important to note that the above arguments and data hold not only for the Amharic person domain, but, according to Schlenker, extend to Engenni\textsuperscript{62}, Aghem\textsuperscript{63}, and Navajo\textsuperscript{64}. Bary & Maier (2003) added Ancient Greek to the list, and Anand & Nevins (2004) Zazaki\textsuperscript{65} and Slave\textsuperscript{66}:

\texttt{(95) \oi \delta \epsilon \iota \nu \sigma \tau \iota \beta \alpha \nu \omega \iota \epsilon \sigma \mu \epsilon \nu \epsilon \iota \zeta \text{the(y).nom.pl part say.past.3.pl that.comp able.nom.pl be.1.pl into}}
\texttt{\tau \iota \chi \omicron \rho \omicron \nu \epsilon \iota \zeta \beta \alpha \lambda \lambda \epsilon \nu \lambda \varepsilon \text{the.land.acc.sg invade.inf}}
\texttt{They replied that they were able to invade this land\textsuperscript{67}}

\begin{quote}
[Ancient Greek. Xenophon, \textit{Anabasis} 5.4.10 (Bary & Maier 2003)]
\end{quote}

\texttt{(96) \text{Hesen.i (mi-ra) va ke ez dwletia}}
\texttt{Hesen.obl (Iobl-to) said that I be rich.pres}}
\texttt{Hesen, said (to me.j) that \{Ij am\} he.i is rich'}}

\begin{quote}
[Zazaki (Anand & Nevins 2004:2)]
\end{quote}

In the temporal domain, Schlenker describes completely analogous shiftability phenomena with the Russian present tense and English \textit{two days ago} and \textit{in two days}, which we will not go into, but while we’re at it, let’s just finish our brief inventory of shifted indexicals, in order to cast further doubt

---

\textsuperscript{62}Nigeria, 20,000 speakers
\textsuperscript{63}Cameroon, 20,000 to 25,000 speakers.
\textsuperscript{64}Athapaskan language spoken by 150,000 people in the US.
\textsuperscript{65}Indo-Iranian language spoken by 2-4 million Kurds in Turkey.
\textsuperscript{66}Athapaskan language spoken in Northwest Territories, Canada (2,600 speaker).
\textsuperscript{67}Interestingly, a very natural translation (e.g. the one given in Xenophon (2005)) uses direct discourse quotation marks, despite the presence of the complementizer \textit{\iota \tau \iota}. It may be possible of course that \textit{\iota \tau \iota} is indeed ambiguous and can mark both indirect and direct quotation. Note further that the continuation of this example would be very hard to translate in English indirect discourse, which may count as an additional argument for treating this as real, direct quotation:

\texttt{[... \epsilon \iota \tau \iota \iota \theta \omicron \iota \tau \omicron \chi \omicron \nu \omicron \upsilon \nu \epsilon \iota \zeta \text{2.pl to the.3.pl for.1.pl to poli.m [[their]]}}
\texttt{[They replied: “We are able to invade this land| of your enemies and ours from the opposite side”]}

We return to the possibility of analyzing indexical shifting as indicating direct discourse in 2.4.3.2.
on the noted objection that these examples do not involve real first person indexicals, but rather some kind of PRO or general anaphor.

In sign languages, a person’s speech (or signs) can be reported by signing the reported utterance, while physically marking that one is playing the role of the reportee. This marking can consist of various things, from mimicking some characteristic aspects of the reportee or her speech, to changes in body posture or eye-gaze. Most indexicals, including first and second person, shift their reference under such a role-shift: 68

\[
(97) \quad \text{George’s facial expression} \quad \begin{array}{c}
\text{George} [\text{George} \mid \text{I win will}] \\
\text{‘George’s like, ‘I’ll win’} [\text{American Sign Language (Lillo-Martin 1995:158)}]
\end{array}
\]

What sets sign languages apart from the other examples discussed so far is that they quite freely allows shifting of the first person outside of speech and attitude reports.

Something similar to role-shifting happens in the spoken language Kwaza 69, which has a speech reporting construction consisting of adding matrix verb person inflexion to a literal copy of the reported utterance: 70

\[
(98) \quad \text{‘she says she is ill’} \quad \text{[Kwaza (van der Voort 2002:312)]}
\]

This same construction, with the shifted first person may be used for animals and even inanimate objects, so it’s really more than just quotation:

\[
(99) \quad \text{‘it is about to run out’ (the gas of the cigarette lighter)} \\
\text{lit: it (says): ‘I’ll be finished!’} \quad \text{[Kwaza (van der Voort 2002:321)]}
\]

As suggested by van der Voort (2002), examples like this one might be explained away by saying that the first person + volitional marking combination has grammaticalized and now denotes something like the verb want/be about to. Otherwise, we might eventually have to include in our analysis indexical shifters beyond the usual attitude operators.

A final piece of data to firmly establish that indexical shifting is not something that can be swept under the carpet by disqualifying dubiously indexical

68 Analyses of role-shifts as monsters have been proposed, recently, by Zucchi (2004) and Quer (2005), more about this below.

69 An isolated language of Brazil with 25 speakers.

70 \text{vol} = \text{volitional marker}, \text{dec} = \text{declarative mood.}
markers from exotic languages, is so-called *free indirect discourse*. Schlenker notes that in free indirect discourse, a literary style mixing the point of view of the character with that of the narrator, some (rather specific) indexical features can be shifted. Consider, for example (100), where the combination of *tomorrow* with the past tense on *was* (both obviously indexical) indicates that they can not both be interpreted as indexical to the same context:

(100) Tomorrow was Monday. Monday, the beginning of another school week! [(Banfield 1982), cited by Schlenker (2003:63)]

It seems that in story telling contexts, indexicals like *tomorrow* and *now* are often interpreted as in the context of the story as experienced from the character’s point of view, while others are interpreted ‘truly’ indexical, i.e. with respect to the context of the narrator (or the reader, depending on the narrative style). In this sense, (100)’s *tomorrow* is a shifted indexical like Amharic *I*. It must however be noted that this literary perspective shifting appears to be a slightly different phenomenon, if only because one of the defining characteristics of free indirect discourse is the absence of a explicitly prefixed reporting verb, i.e. the shifts occur freely in main clauses. Though Schlenker (2004a) has abandoned the view, advocated in his (1999) and Sharvit (2004), that free indirect discourse constitutes a genuine monster, such a view would still be preferable over an analysis that stipulates that somehow this *tomorrow* is not the indexical it is in ordinary spoken English.

### 2.4.3.2 Shifty indexicals as quotation

A second maneuver that can be performed to avoid monsters is to claim that we are really dealing with quotation here. The argument originates with Kaplan (1989:510-1) who, after formulating the Prohibition on Monsters, observes that English does seem to have a way to shift the reference of indexicals, viz. by quoting them:

(101) Otto said “I am a fool.” [(Kaplan 1989:511)]

He goes on to claim that such direct discourse-type quotations *mention* rather than *use* the quoted expressions. This means that the quoted first person pronoun in (101) is not used to refer to any individual whatsoever, rather the whole quoted phrase refers to an utterance. On this view, quotation marks turn expressions of the object language into terms referring to those expressions, just like we sometimes use the Greek letter $\varphi$ to refer to a formula, or a bracketed natural number to refer to an example sentence. In all those cases we do not *use* the intended expressions, but make claims *about*
them, which is what makes Kaplan view all statements exemplified below as metalinguistic:

\[(102) \quad \begin{align*}
\text{a. } & \text{A formula } \varphi \text{ is valid iff it is true in all models} \\
\text{b. } & \text{This is illustrated in sentence (37c)} \\
\text{c. } & \text{Otto’s favorite one letter word is } I
\end{align*}\]

Adhering to the logician’s strict use/mention distinction, Kaplan considers any form of metalinguistic reference to fall outside the scope of semantic theory, perhaps for fear of infesting the logic with paradoxical self-reference. But even if we were to widen the scope of semantic theory so as to accommodate (102)’s perfectly fine specimens of naturally occurring, spoken and written, English, by somehow adding devices to refer to expressions of the object language to that same object language, we wouldn’t need monsters: from the fact that the occurrences of \(I\) in (101) and (102c) refer to linguistic entities rather than persons we can already derive the apparent indexical shift in the latter.

Along these lines, direct discourse in English is rendered harmless for Kaplanian theory. Now, can we disqualify all the above shifting data by analyzing the reporting constructions as quotation? Well, putting aside the temporal shifts, and the free indirect discourse, which was a bit odd anyway, this seems a prudent course of action. At this point it does indeed seem quite feasible to analyze our examples as involving optional, hidden quotation marks: John says: “I am a hero” for (93), They said: “We are able” for (95), Hesen said “I am rich” for (96), and, finally, note that Lillo-Martin’s (1995) translation of the signed example already uses a direct discourse construction with quotation marks. The observed ambiguity between a shifted and a non-shifted, Kaplanian reading (cf. fn. 61 and (96)) shows that the quotation marks are not realized on the surface and the hearer may choose to interpret the reports with or without them. This may not be the prettiest solution (hidden operator, ambiguity) but in its defense it must be noted that quotation marks are often not that clearly pronounced in spoken English either, at least not as separate morphemes.\(^{71}\)

\(^{71}\)It does seem that often a very specific intonation pattern marks quotation in English, perhaps following Potts’ hypothesis:

**Lexicalization hypothesis** In quotation, each prosodic word projects to its own intonational phrase with a rise-fall-rise contour. 

\[([ \text{Potts 2005:8} ])\]

The extent to which this covers all forms of quotation (from pure mention to scare quoting) and successfully disambiguates, definitely requires further research:
Anand & Nevins’s (2004) analysis of Zazaki shifting lends some prima facie credibility to the quotation analysis. In Zazaki, indexicals obey what they call the ShiftTogether constraint, which says that indexicals may (or may not) shift under an attitude operator, but they cannot shift independently. They provide data like (103), to show, convincingly, that once an indexical shifts, all the other indexicals in the same report shift too. Consider:

(103) Vizeri Rojda Bill -ra va ke ez to -ra miradiša
    yesterday Rojda Bill -to said that I you -to be angry.pres
    ‘Yesterday Rojda said to Bill that she was angry at him’
    [Zazaki (Anand & Nevins 2004:4)]

With two embedded indexicals (disregarding the present tense on the verb), and the possibility of shifting or not, we might expect three more readings beside the one given in the translation. As it turns out, the sentence has only two: either both indexicals or neither is shifted. For ease of presentation, imagine me uttering (103) to Hesen (in his native language). The possible readings are:

(104) Rojda to Bill:
    a. “I am angry at you”
    b. “Emar is angry at Hesen”
    c. *“Emar is angry at you”
    d. *“I am angry at Hesen”

For completeness, note that the two readings of the Zazaki report correspond exactly to what we get on a direct vs. indirect discourse ambiguity, as predicted by a quotational analysis:

(105) a. Yesterday Rojda said to Bill, “I am angry at you”  [≈(104a)]
    b. Yesterday Rojda said to Bill that I am angry at you  [≈(104b)]

Anand & Nevins (2004) formalize the ShiftTogether constraint by postulating an optional indexical shift operator, introduced by the attitude verb, together with an implicit, yet in my view quite bold, assumption that indexicals have to be interpreted in situ (i.e. they can never escape the scope of the nearest such operator by means of, say, presupposition projection or covert movements).\footnote{My own approach in chapter 3 is one that treats indexicals as presuppositions that prefer to take widest possible scope. This follows earlier DRT analyses of referential expressions: Geurts (1997) for proper names, and Zeevat (1999) for indexicals.} The operator-theoretic approach faithfully captures ShiftTogether and makes an interesting further prediction, viz. that “[i]n cases
of multiple embedding, shifting in the intermediate clause prevents a lower indexical from being interpreted indexical to the matrix context” (p.16), which they show to be borne out by the data as well.

Now, is ShiftTogether a universal constraint, and if so, is there an empirical difference between the operator-theoretic and the quotational analyses? As for the first question, it seems the answer must be no. For consider:

\[
\text{IX}a \text{ MADRID}_m \text{ MOMENT}^5 \text{ JOAN}_i
\]
\[
\text{place}_a \text{ Madrid at time t John}
\]
\[
\text{THINK IX-1, STUDY FINISH HERE}_b
\]
\[
\text{[\text{\text{–}John think I study finish here}]
\text{‘When he was in Madrid, John thought he would finish his studies here (in Barcelona)’ [Catalan Sign Language (Quer 2005)]]}
\]

In this sign language example, the first person is shifted, but the locative indexical here gets a Kaplanian interpretation. Even within the person domain we can get similar mixes:

\[
\text{Simon rásereyineht’ù hadi}
\]
\[
\text{Simon 2.sg-hit-1.sg say.3.sg}
\]
\[
\text{‘Simon said that you hit him’ [Slave (Anand & Nevins 2004:5)]}
\]

This last one may not be too clear, because the English sentence is compatible with a fully shifted reading of the original as well. To highlight the mixed reading, imagine I utter (107) to Ellsworth, because I’ve heard that Simon has been going around telling people, “Ellsworth hit me”, behind her back. In this case the full shift reading, Simon said, “you hit me”, is out, but the Slave example should still be true.\(^3\)

In light of the Slave data in particular, Anand & Nevins propose a somewhat weaker version of the ShiftTogether constraint which allows for language variation with respect to the indexical domain. In other words, for Zazaki, there’s a ShiftTogether for all indexicals, whereas in Slave it holds only for first person indexicals, and in Catalan Sign Language perhaps for all but locative and temporal indexicals. Consequently, the operator-theoretic approach must be adapted to affect only a particular group of indexicals. Note that, even combined with the optionality of shifting, the predictions are still non-trivial, because (i) indexicals of the exact same type in the same clause must still always shift (or not shift) together, and (ii), in case of

\(^3\)Anand & Nevins’ discussion leaves this implicit, but that’s what their gloss and analysis predict.
multiple embeddings, a shift in some shallow embedding prevents deeper indexicals (of the given type) from being indexical to the main context. Anand & Nevins claim that this holds for Slave and Zazaki. For other languages rigorous testing has yet to be done. A tentative counterexample to (i) might be Amharic: 74

(108)  Abbottazzaaññi alɔ 
1.sg-will-not-obey-1.sg 3.sg.m-past-say 'He, said he, would not obey me,’ [Amharic (Leslau 1995:779)]

In this example the embedded clause contains two first person markers, only one of which is interpreted as shifted. But, before we throw out the ShiftTogether constraint too hastily, note that further research may be necessary to confirm the Amharic judgment, as it has been noted that the example is “not very solid” (Schlenker p.c. to von Stechow, who uses this example to show that Amharic shifting is not simply reducable to direct discourse). We must therefore leave the matter of the alleged universality of ShiftTogether for future research.

So, a weakened version of ShiftTogether may hold, but with the restriction to a particular domain of indexicality we have immediately lost the close parallel with the direct discourse analysis which did equivalent predictions for the Zazaki examples considered so far, including the one with the multiple embeddings. But, obviously, the mixed indexicality examples in (106), (107) and (108) cannot be analyzed by quoting the entire report complement. Thus, as to the second question—is ShiftTogether equivalent to a direct discourse analysis?—the answer must be no, quotation is much stronger, because it necessarily affects every term in its scope, not just indexicals. Counterexamples to the quotation analysis are therefore provided by mixed cases that retain some feature as if they were in direct discourse, while adapting others to conform to the format of indirect discourse. We’ve already seen some, but these mixes need not be confined to indexical features:

(109)  eli δε δη elpe [...] πεμψα δε και προκαταληψομενους one prtprt said [...] to send.infprt and occupy.ptc.acc.pl 
та ᵣαρα, ᵥπως μη φυλασσαι μητε Κυρος the heights.acc, so that not forestall.past.3.pl neither Cyrus.nom

74 This example is cited by von Stechow (2001:5) and (2002:9), as well as Anand & Nevins (2004:15), though Schlenker (2003) doesn’t mention it, presumably because it “is not very solid” and there are better investigated alternatives to show that indexical shifts are not quotation, like (116) (Schlenker p.c. to von Stechow)
Chapter 2. Attitude reports

In (109), the main verb of the reported clause is an infinitive, πέρψα, which indicates that the report is indirect, for in direct discourse the relevant commander must have used a finite form of the verb *come*, probably 1st person plural like the shifted/quoted εχομεν in the relative clause. On a quotation analysis, this doesn’t add up: apparently, the report is indirect except for (part of) the relative clause. In English, it has been argued that such intrasentential mixes of direct and indirect discourse make little sense, so we can give up the quotation analysis:

(110) ??one man in particular proposed [...] to send a force [...] in order that neither Cyrus nor the Cilicians, of whom “we have seized many people and things as plunder,” should forestall them

On the other hand, a quick look at your local newspaper suffices to show that mid-sentential mixes of indirect and direct discourse are ubiquitous in written language, and indeed the translation given in Xenophon (2005) relies on just such a mix:

(111) one man in particular proposed [...] to send a force [...] in order that neither Cyrus nor the Cilicians should forestall them—“and we have in our possession,” he said, “many of these Cilicians and much of their property that we have seized as plunder.”

Though this particular sentence may be a rather special case, as it’s introduced by a dash and starts (almost) at a sentence boundary, the recent literature on subclausal quotation shows that it’s quite a powerful tool (Cappelen & Lepore 1997; Recanati 2001; Potts 2004a; de Brabanter 2005; Geurts & Maier 2005). Perhaps, then, *mixed quotation*, as Cappelen & Lepore (1997) have termed the mix of direct and indirect discourse paradigmatically exemplified by (112), is exactly what we need to explain away these alleged
indexical shifting monsters?

(112) Quine says that quotation “...has a certain anomalous feature”

[(Davidson 1979:30)]

Note for instance that mixed quotations in English appear to shift indexicals:75

(113) a. Their accord on this issue, he said, has proved “quite a surprise to both of us.”

[New York Times, (Cappelen & Lepore 1997)]

b. Pascal suspected that the mercury was really supported by the “weight and pressure of the air because I consider them only as a particular case of a universal principle concerning the equilibriums of fluids.”

[(Hall 1981:252)]

c. Niet alleen de RAI, ook het parlement, nee, heel Italië is Not just the RAI, also the parliament, no, all of Italy is trots op “onze meiden” in Irak

proud of “our girls” in Iraq

[Dutch, Volkskrant, 19-4-’03]

d. Hij vindt iemand stoer als ‘hij is zoals mij’

He thinks someone cool if he is like me.acc

‘He thinks somebody’s cool if “he is like me’”

[Dutch, Contrast, 20-12-’02]

e. Naturally John Lennon was expelled and sent to art school, “so I can fail there as well”.

[Guardian, 2-11-’00, cited by de Brabanter (2005:2)]

f. [The] commercial for a popular pain reliever which is endorsed by an actor who is “not a doctor, but I play one on TV”

[itre.cis.upenn.edu/~myl/languagelog/archives/002541.html]

Apparently, subclausal quotation is not only ubiquitous in English, it’s also powerful enough to shift some indexicals in a report, while leaving other parts of the same clause in indirect form. Quotation might go a long way in explaining indexical shifts after all.76 It remains unclear whether we can adequately capture all the apparent direct/indirect mixes encountered above

75I have discussed the matter of indexical shift in mixed quotation elsewhere, along with some of the examples in (113), e.g. in (Maier 2003)

76It is not clear what are the restriction on mixed quotation as it is still a relatively new field of study. Some authors argue, for instance, that subclausal quotation can only apply to syntactic constituents (Geurts & Maier 2005), while others say this is too strict (Cumming 2005). (We’ve already seen apparent counterexamples to the constituent hypothesis in (113b) and (113f), though there may be room for a fix along the lines of breaking down the larger quotation into two parts, at the clausal boundary indicated by the conjunction.)
by mixed quotations. Note that the mechanisms of mixed quoting must be quite flexible indeed. I offer some more empirical evidence to cast further doubt on the quotation analysis. These involve NPI licensing, indirect questions, and extraction. Besides, there’s also a purely theoretical complication in using mixed quotation, but data first.

So-called negative polarity items (NPIs) are terms that can occur only in negative, i.e. downward entailing positions in a sentence, like, say, under the scope of a negation, or in the antecedent of a conditional. In Zazaki kes is such an NPI—as is, by the way, its translation, anyone, in English. As in English, a negative marker in the main clause may license this NPI in an indirect discourse report complement, but interestingly it does the same if that complement contains shifted indexicals:

(114) Rojda ne va kə mi kes paci kərd
Rojda not said that 1.erg anyone.NPI kiss did
‘Rojda didn’t say that she kissed anyone’

[Zazaki, (Anand & Nevins 2004:3)]

Obviously then, we’re not dealing with direct discourse for the complement on its own is ungrammatical, as is its English counterpart:

(115) *Rojda didn’t say: “I kissed anyone”

The translation obviously doesn’t allow mixed quotation that quotes the first person while leaving the NPI object alone, but perhaps that’s just a typological difference in the rules governing mixed quotation.

The same kind of observation can be made with respect to belief embedded wh-question formation: In English, the required wh-extraction is possible in indirect, but not direct discourse; in Amharic it’s possible in reports with shifted indexicals:

(116) min amt’ -a ind -al -ṇṇ al -səmma -hu
what bring.imper -2.m comp -say -3.m -1.sg.obj neg -hear -1.sg
-mm
-neg
‘I didn’t hear what he told me to bring’

[Amharic, (Leslau 1995:779)]

As for terminology, we say that the wh-phrase what is ‘(base-)generated’ as the object argument of the embedded imperative bring, and then, in the English translation at least, ‘extracted’, i.e. moved out of its clause to surface somewhere else, leaving but an ‘invisible trace’ at the object position it started at.
To see what this example is supposed to show, let me sketch you a scenario in which it could occur:

(117) In the distance, a rather agitated fellow is yelling to me: “Bring my grrm...!” Since I can’t make out what he wants me to bring, I turn to a nearby Ethiopian and utter (116), i.e.

a. I didn’t hear what he told me to bring
b. *I didn’t hear that he said to me “bring what!”
c. *I didn’t hear what he told me to “bring!”

The shifted indexical is the second person feature on the embedded bring. In addition, the fact that this verb is in imperative mood also points to a direct discourse-like quotation. But given the scenario in (117), the embedded object, what, is certainly not quoted (cf. (117b)): it is what makes the report an indirect question, as in (117a). So, if we really want to maintain it’s quotation, we are forced to accept some rather shady mix quotes, as in (117c).

The same point can be made for (other types of) extraction from a report complement. As a final example, consider:

(118) čćęčče [kè Ḗseni va mi paci kërda] rindèka

girl that Hesen said I kiss did be pretty.pres’
the girl that Hesen said he kissed is pretty’

[Zazaki, (Anand & Nevins 2004:3)]

According to Anand & Nevins (2004:3), this extraction of the girl from the object argument slot of kiss “is illicit in bona fide cases of direct discourse”. Note the cautious use of bona fide, perhaps in order to leave room for a weakened, mixed quotational analysis.77

77The proposal under investigation here (i.e. that shifted indexicals are mixed quoted) bears some striking outward resemblance to Schlenker’s (1998) idea that shifted indexicals are indirectly quoted. In Schlenker’s terminology, a report is an indirect quotation iff its complement’s LF corresponds to the reported utterance’s. Note that Schlenker’s LFs are highly structured semantic entities, think syntactic trees with semantic leaves, making indirect quotation something in between direct quotation and indirect speech, hence the name. He successfully employs indirect quotation to account for intraclausal combinations of shifted indexicals with extraction, like (118). However, since his indirect quotation operator ‘QUOT’ takes scope over the whole embedded clause, examples of mixed indexicality, like our (106), require something extra, viz. that “constituents of a quoted LF may be scoped out of QUOT and interpreted à la Kaplan [(1969)]” (p.3). This is no longer compatible with our tentative mixed quote analysis, which does not require scooping of indexicals, but instead adjusts the scope of the quotation. I intend to explore the various ways one can go about implementing a ‘quotational analysis’ of shifted indexicality in future research.
This concludes a brief overview of the shifty data. It shows that the pseudo-indexical/anaphoric and direct discourse analyses are too simple. The status of the ShiftTogether constraint is still an open question (along with the possibility of an operator-theoretic approach), as is the possibility of a weakly (mixed) quotational analysis. Before going on with Schlenker’s alternative analysis in terms of context quantification, let me close this quotation section with a theoretical issue concerning mixed quotation.

As is well-known in the quotation literature, mixed quotation is logically quite distinct from direct discourse. The first one to pinpoint the issue was Davidson:

Quine says that quotation “...has a certain anomalous feature”

Are the quoted words used or mentioned? Obviously mentioned since the words are Quine’s own, and I want to mark the fact. But equally obvious is the fact that the words are used; if they were not, what follows the word ‘quotation’ would be a singular term, and this cannot be if I have produced a grammatical sentence.

In other words, mixed quoted phrases cannot be analyzed as mere mention, for that would turn the phrase into a referring term, which, given the position of the quotation marks, often makes little sense, grammatically speaking.

Additional evidence for what I call the use hypothesis for mixed quotation comes from anaphora across quotation marks.\(^78\)

David Penberthy […] has said that his paper’s stories about Mr Brogden’s behaviour, quote: “Wouldn’t be appearing, if there weren’t people, inside the Liberal Party who were trying to get them out,” unquote. Who were these people inside the Liberal Party?

This example shows that a quoted expression can still contribute a ‘dis- course referent’ to the common ground, which indicates that, at some level, these expressions are ordinarily interpreted, as if used, in addition to any mention-based meaning they may have.\(^79\) Another piece of data supporting the use hypothesis concerns the occasional occurrence of unshifted indexicals in mixed quotation:

\(^78\) There’s a purely technical reason for the fact that (119) and (120) below both contain the, not all that common, phonetically spelled out quote/unquote construction: there’s no way to Google for (most) punctuation marks.

\(^79\) Interestingly, the same thing holds for full, direct quotations:
2.4 Anti-reductionism and pure de se separatism

(120) And I even pissed off the youngest one so much that he told me to quote unquote ‘stick a lamp up my ass’.

[http://www.fansfromoz.com/lowie.html, 21-11-'05]

We see here an unambiguously Kaplanian interpretation of my, referring to the actual speaker, who was probably addressed in second person originally. Apparently, the pronoun is not mentioned or it would have retained its direct discourse form. In connection with the indexical shifts in mixed quotation observed earlier, (113), note that it’s a completely open question why some mixed quoted indexicals shift, while others do not.

As Davidson also noted, we’re not just dealing with ordinary use either, mixed quotation combines elements of both modes of reference. In the recent literature there has been a lot of discussion about the (relative) status of the use and mention components of a mixed quotation’s ‘meaning’: some say the mention part belongs to a fully independent layer of content (Potts 2004a); others that it is ‘merely’ pragmatic—with truth-conditional impact (Recanati 2001; Predelli 2003), or without (Geurts 2001).\(^8\)

In any case mixed quoted expressions are not just mentioned, but used. Now remember that the whole point of the quotational enterprise was to reduce shiftiness to mention. The reasoning was as follows: if, as in pure mention or direct discourse, an indexical refers to a linguistic item instead of a contextually supplied individual, it is hardly surprising that it violates Principle 2. Moreover, if, as in direct discourse and mixed quotations, the words mentioned are specified as belonging to the reportee’s original utterance, we get the fact that quoted indexicals appear to be interpreted from that reportee’s point of view/context for free. Now that it has become clear that in mixed quoting we’re really using the so quoted indexicals, we are at a loss to explain why we can get shifts. The use hypothesis for mixed quoting, plus the apparent optionality of indexical shifts therein, make the situation remarkably similar to, and at least as complex and ill-understood as, the cross-linguistic data we were trying to analyze with it. Perhaps then it makes more sense to turn things around and start from an account of shifting in indirect discourse, and then reduce mixed quotation to that. The next

\(^{8}\)Geurts & Maier (2005) argue for a ‘one-dimensional’ analysis, unifying the two levels into one: the deferred use of the mentioned constituent.
section will provide just such an analysis of indirect discourse: Schlenker’s analysis of shifty indexicals, formulated in terms of a monstrous quantification over contexts, independent of quotation.

2.4.3.3 Schlenker’s Monsters

Schlenker (2003) proposes an alternative account of the data with logophors, PRO, and shifted indexicals. It treats these elements as indexicals that can pick up a referent from a reported speech/attitude context, which in turn is made available by the attitude verb. His analysis requires significant extensions to the logical language, first of all, he moves from an intensional logic to a many-sorted extensional system, replacing modal operators with quantification over worlds and times, as exemplified below:

\[
\begin{align*}
\Box \neg \exists x [\text{fish}(x) \land \text{walk}(x)] \rangle_{w,t}^f \\
\leadsto [\forall w' \forall t' \neg \exists x [\text{fish}(x, w', t') \land \text{walk}(x, w', t')] \rangle_{w \mapsto w', t \mapsto t}'^f \\
\end{align*}
\]

Note how the assignment function now takes over the role of the intensional interpretation parameters (which in the new system are represented by free variables), and how all predicates suddenly require two extra argument positions. The result of this move is a seemingly more complex and powerful system but, as Schlenker points out, at least this much expressiveness is needed on independent linguistic grounds.

To handle indexicals we already went beyond intensional logic and added a second interpretation parameter, the context. We could just leave that one as is, and mimic Kaplan’s logic:

\[
\begin{align*}
\Box \text{walk}(i) \rangle_{w,t}^{f,c} \\
\leadsto [\forall w' [\text{walk}(i, w', n)] \rangle_{w \mapsto w', t \mapsto t, c}^{f,c} \\
\end{align*}
\]

Note how, as a bonus, we have a more unified representation of indexicals in different domains, i.e. we have constants \(a\)ctually and \(n\)ow, like \(i\), instead of operators \(a\) and \(n\). And since Kaplan’s prohibition of monsters required that belief reporting is an intensional operator, like \(\Box\), it too becomes a world quantifier, albeit a parametrized one:

\[
\begin{align*}
\text{BEL}_{w,t}^{\text{that}} \text{walk}(i) \rangle_{w,t}^{f,c} \\
\leadsto [\text{BEL}_{w,t}^{\text{that}} w' [\text{walk}(i, w', n)] \rangle_{w \mapsto w', t \mapsto t, c}^{f,w}[w \mapsto t]}^{f,c} \\
\end{align*}
\]

The semantics of this quantifier can be copied mutatis mutandis from the one discussed in 2.3.4, but let’s leave our extensional emulation of Kaplan
and move on to the more advanced system that can cope with the data of this section.

As previously announced, the logic will be monstrous in the sense that its attitude operators are really quantifiers over contexts. A last addition is therefore a fourth quantification type, the context. Truth will always have to be defined relative to a context, outside the object language, but with the addition of context variables in the language we can specify which contexts a given indexical depends on. To achieve this we turn our indexical constants into function symbols, mapping context variables onto individuals. The indexical now or a present tense marker, for instance, are represented at lf as \( t(c) \), consisting of the indexical ‘now’ function mapping each context onto its time coordinate, and a context variable. As a notational shorthand we’ll always just subscript the context argument to an indexical. The first person thus comes out as \( a_c \), where \( \llbracket a_c \rrbracket^{c \rightarrow c} = a_c \) = the agent/center/speaker of context \( c \).

Let’s go through an example to make things clearer:

\[
\begin{align*}
(124) \quad & \llbracket \text{BEL}_{e}^{\text{THAT}} \text{walk}(i) \rrbracket^{f,c} \\
\sim & \llbracket \text{BEL}_{e,w,t}^{\text{THAT}} c' [\text{walk}(a_c, w_{c'}, t_c)] \rrbracket^{f[w \rightarrow w], [t \rightarrow t], [c \rightarrow c]} \end{align*}
\]

Using the notion of a belief context, \( c \in B_{e}^{C} \), as described for example in 1.2.3 and 1.B.4, the semantics of this monster is as follows:

\[
(125) \quad \llbracket \text{BEL}_{e,w,t}^{\text{THAT}} c' [\text{walk}(a_c, w_{c'}, t_c)] \rrbracket^{f[w \rightarrow w], [t \rightarrow t], [c \rightarrow c]} = 1 \quad \text{iff} \quad \text{for all } c' \in B_{e}^{C}(\text{ellsworth}, i) : \langle a_{c'}, w_{c'}, t_{c'} \rangle \in \llbracket \text{walk} \rrbracket
\]

Assuming the utterance context, \( c \), is the actual context centered around me, while I’m writing this, this means (124) is true iff in every context compatible with Ellsworth’s belief the world is such that Emar walks October 8, 2006.

With our new machinery, we can say that some indexical constants, like English I and now, as demonstrated above, always take the main context \( c \) as their argument, while shiftable elements like Amharic I can take either that, or the one introduced by the attitude verb. In that last case we get:

\[
(126) \quad \text{John thinks I am a hero} \quad \text{[Pseudo-Amharic, \approx(93)]} \\
\Downarrow \quad \llbracket \text{BEL}_{j,w,t}^{\text{THAT}} c' [\text{hero}(a_{c'}, w_{c'}, t_{c'})] \rrbracket^{c} = 1 \quad \text{iff} \quad \llbracket \text{BEL}_{j,w,t}^{\text{THAT}} c' [\text{hero}(a_{c'}, w_{c'}, t_{c'})] \rrbracket^{f[w \rightarrow w], [t \rightarrow t], [c \rightarrow c]} = 1
\]
Chapter 2. Attitude reports

iff for all $c' \in \text{Bel}^C(john, \langle w_c, t_c \rangle)$: $(a_{c'}, w_{c'}, t_{c'}) \in [\text{hero}]

That is, (126) is true iff in all of John’s actual belief contexts the center is a hero. This is exactly the pure de se belief that in English we would normally report with John thinks he’s a hero, but that in Pseudo-Amharic would be spelled out as (126). This immediately leads to an interesting prediction: unlike the English gloss with the coreferential pronoun, we predict that the Amharic shifted reading is unambiguously pure de se. The quotational account shares this prediction, for it requires directly that John used the first person pronoun, and so does the anaphoric account, which treats shifted I as a he* (which is defined as allowing only pure de se readings). Unfortunately, as far as I know, this prediction has not been tested (e.g. by using a pants-on-fire scenario, as Kusumoto (1998) did with a Bafut logophor, here in (55b)). For completeness, here’s a representation of the de re/de se ambiguous English coreferential pronoun report, pure de se reading first:

$$\ell^{C}$$

Ellsworth believes she is going to win

$$\ell^{C}$$

A. $\ell^{C}$

$$\ell^{C}$$

Before we go on, it should be noted that many of the shifty data are tested (or even known to hold) only with indirect speech rather than belief reports (cf. the genuine Amharic (93) that gave rise this Pseudo-Amharic adaptation), but with minor modifications, essentially the same story can be told for speech as for any other type of attitude report, we just need to define $\text{Say}^C$, the set of contexts compatible with what a person said:

$$\ell^{C}$$

In other words, $\text{Say}^C(a,i)$ is simply the diagonal of the utterance of $a$ in $i$. Obviously this infests the semantics of speech reporting with a very high degree of omniscience/closure, but, as argued before, this does not concern us here.

Some examples from 2.4.3.1 and 2.4.3.2:

$$\ell^{C}$$

They replied that they were able [to invade this land]’
2.4 Anti-reductionism and pure de se separatism

\[ \exists t' < t [ \text{say}^{\text{that}}_{x,w,t'} c'[\text{able}(a_{c'} + Y, w_{c'}, t_{c'})]] \]

(130)  Ιx a MADRID\_n \_MOMENT\_i JOAN\_i
place\_a Madrid at time \_t John THINK IX\_i STUDY FINISH HERE\_b RS\_i

[\[ \text{say}_X \text{that}_Y \text{X}, w, t' c' \\text{able}(a_{c'} + Y, w_{c'}, t_{c'}) \]]

In all of these the indexicals, shifted or not are represented as logical indexicals, i.e. as individual constants/functions that get their referent from a context. Some indexicals, like I and now in (124) and here in (130), take the actual context, c, which means they get a Kaplanian interpretation. Others may pick up the quantified variable \( c' \) introduced by the report, so that they receive a shifted interpretation, e.g. the first persons and the, not so directly surface-realized, tense and mood arguments of the embedded predicates in (129)-(131).

You’ve been promised a unified account of shifts with PRO and logophors, so let’s now turn to those other data, from 2.4.1. Of course, the way to go is pretty obvious: we’ve already seen how to write purely de se logical forms, apparently we just need to stipulate that indirect reflexives, logophors and PROs are always represented as shifted first person indexicals, i.e. a\( c' \).

(132) ??Philip believes that he himself is dead

\[ \exists t' < t [ \text{say}^{\text{that}}_{r,x,t'} c'[\text{dead}(a_{c'}, w_{c'}, t_{c'})]] \]

---

\(^{81}\) The first person plural is simplifyingly represented as a\( c + Y \), i.e. the context’s center plus some other individuals left unspecified. The free plural variable Y should be thought of as a presupposition or anaphor, as it can be bound to any salient set of individuals available in the discourse or the wider context. In this case, the, in part pragmatically determined, assignment function should map Y onto the denotation of X minus the speaker, a\( c' \).

\(^{82}\) \( p_c \) denotes the place of context c, i.e. the place where a\( c \) is at \( t_c \) in \( w_c \).

\(^{83}\) In the case of second person de se reports, the logophor or PRO is represented as h\( c' \) (= the addressee/hearer of context c’).
160  

Chapter 2. Attitude reports

(133) kofi be yè -dzo  
Kofi say LOG -leave  
‘Kofi said he’d left’  

[≈(54a)]  

(134) Ellsworth believes PRO to be on the winning side.  

[≈(59b)]  

In a sense, we’re doing the exact opposite of the anaphoric proposal denounced in 2.4.3.1: we reduce the observed pure de se reporting through PRO and logophors to monstrously shifted indexicality, instead of analyzing shiftability of indexicals as a special case of Chierchia-style logophoric binding (by intervening property abstractors).

Having gone through a couple of examples above, note how we’re led to a three part typology of indexicals: there are rigid ones, shiftables, and obligatorily shifted ones, exemplified in the person domain by English I, Amharic I and PRO, respectively. The lexical stipulations, preliminarily cashed out in terms of the features ±actual and ±c(ontextual), that say which indexical form belongs to which category Schlenker calls the filtering mechanism. Part of Schlenker’s paper is devoted to a derivation of this filtering mechanism (using a theory of presuppositions).

Besides filtering, which is necessarily stipulative to a degree, Schlenker’s full semantics requires a so-called agreement mechanism to explain why some surface features of indexicals in reports appear to be left uninterpreted. One example would be the third person feature of the embedded subject in a ‘co-referential’ third person report, like (127). I put scare quotes on coreferential because in the proposed pure de se If, (127a), the second pronoun is really represented as independent of the matrix clause. Rather, the embedded she is represented as a_c', denoting the center of a context, i.e. a semantically first, rather than third person. Reasoning in the opposite direction, from deep to surface structure: something is needed to explain why the pure de se If’s a_c' can surface either as a third person pronoun (in English, (127)), a third person logophor (Ewe, (133)), a first person pronoun (Amharic, (126)), or a (silent) PRO (English, (134)). Amharic represents the most natural grammaticalization; simply analyzing the Amharic first person as carrying the features +C and 1ST does the trick.

Leaving the pronominal realizations for what they are, Schlenker focuses on PRO, which, he shows, even though it’s invisible and semantically first person, actually carries the person, gender and number features of the matrix subject. That it has these features is easily brought out by making it bind a reflexive:
2.4 Anti-reductionism and pure de se separatism

(135) John hopes PRO to buy \( \{ * \text{myself} \} \) a car

\[ \text{hope}_{j,w,t}^\text{that} c'[\text{buy\_car\_for}(a'_c, a'_c)]^{84} \]

In (135), PRO binds the reflexive, so their features must match, from which it follows that PRO is not morphosyntactically first person. Moreover, if we ensure that in all hope alternatives \( c' \) the center buying the car has become female, we still get a masculine reflexive and PRO, provided that the main clause subject, John, is a man in the actual context:

(136) John, a transsexual, hopes to become a woman and then \([\text{he hopes to}]\) buy \( \{ * \text{herself} \} \) a car

\[ \text{hope}_{j,w,t}^\text{that} c'[\text{woman}(a'_c) \land \text{buy\_car\_for}(a'_c, a'_c)] \]

Conclusion: PRO is a third person, masculine pronoun that may refer to a female center, meaning that its hidden features are somehow deleted at LF.

In the temporal domain, we find a completely parallel situation, already extensively documented, e.g. for Latin, as part the 'consecutio temporum' (sequence of tense) rules:

(137) Ignoravi quid ageres

\[ \text{not know.1.sg.ind.past(perf)} \land \text{what.2.sg.subj.past(impf)} \]

'I didn’t know what you were doing'


\[ \exists t' < t[\neg \text{KNOW}_{a'_c,w,c'}^{a_c,w,t_c} c'[\text{do}(h_c, ?, w_c, t_c)]] \]

In (137) the embedded past tense contributes a \( t_c' \) to the LF,\(^{85}\) reflecting the fact that the embedded action is interpreted as simultaneous\(^{86}\) with the not knowing. We might also describe this by saying that the formal past tense feature of the embedded verb is deleted in the derivation of LF. As the translation in (137) shows, the same thing happens in English. Now, there are some alternative analyses of the simultaneity of past-under-past (e.g. Enç (1987)), but we will not go into those, or their limitations. Instead let’s follow Schlenker (1999)\(^{87}\) who stipulates sequence of tense rules that say that

\(^{84}\)We occasionally suppress some irrelevant, obvious arguments, like in this case \( w_c', t_c' \).

\(^{85}\)Note that the subjunctive mood can be seen as a realization of a shifted \( w_c' \) argument. In fact, Schlenker argues that the subjunctive is a logophor in the possible worlds domain.

\(^{86}\)Or, perhaps, ‘pseudo-simultaneous’, in reference to our earlier remark on coreferentiality in agreeing third person reports on p. 160.

\(^{87}\)In later work, Schlenker also tries to formulate ways to get around sequence of tense, and of person stipulations.
under certain conditions, embedded tense inflections are not semantically interpreted, thus allowing simultaneous, pure de se interpretations of e.g. past-under-past constructions. I shall not formulate here what these conditions are, for we’d better return to the person domain, where things are more manageable.

Acknowledging the parallel between de me readings of ‘third-under-third’, and de nunc readings of past-under-past constructions, Schlenker proposes the addition of a sequence of person rule to semantically delete person and gender features on embedded PROs and other pronouns. But, the question remains, when do we delete features, and when do we interpret them? Answer: embedded features may be deleted at lf when they agree on the surface with the matrix subject of the report in question. Or, formulating minimalistically, i.e. the other way around, in the morphological spell-out process, embedded (pseudo-coreferential) pure de se pronouns inherit the person, gender and number features from the main clause’s subject, to achieve morphological agreement. The idea that indexical features can get lost under morphological agreement is not that new, so let’s trace its history to see it in action, and hopefully to get some independent motivation for it. This will conclude our discussion of Schlenker’s (1999; 2003) semantics of attitude reports.

2.4.3.4 Binding and feature deletion under agreement

Starting from Chierchia’s idea of a complement phrase introducing a λ as property abstractor (2.4.1.3, 2.4.3.1), Heim (1994) describes how embedded (indexical) pronouns may be represented as variables bound by such an attitudinal abstractor at lf. For Heim, the possibility of bound variable interpretations is governed by an agreement constraint: a person pronoun in a report complement may be represented as a bound variable iff its person, number and gender features match those of the main clause subject. For example:

(138) I think I’ll win
\[ \triangleright \text{BEL}_1^r \lambda x [\text{win}(x)] \]

The embedded pronoun in (138) has the same features as the report’s subject and therefore it may be bound by the intervening λ. The semantic effect of this binding is that we get a pure de se reading, the speaker self-ascribes the property of winning. Note how, in the process, the indexical has lost all its referential features until only a bare variable remains at lf, an aspect of the current analysis that Schlenker’s monstrous theory of shifted indexicality does not adopt.
Obviously, the agreement constraint is not vacuous, it’s needed to avoid unwanted pure *de se* readings in cases of feature mismatch:

\[(139)\] Ellsworth thinks I’ll win
\[\triangleright\ \text{BEL}^*_e \lambda x [\text{win}(i)]\]

Further, the optionality of binding under agreement is necessary to account for the so-called ‘*de re*/*de se* ambiguity’ observed time and again for third person reports:

\[(140)\] Ellsworth believes she’s going to win
\[\begin{align*}
a. & \triangleright\ \text{BEL}^*_e \lambda x [\text{win}(x)] \\
b. & \triangleright\ \text{BEL}^*_e \lambda x [\text{win}(e)]
\end{align*}\]

Completely in line with Chierchia’s ambiguity thesis, \((140)\) is ambiguous between \((140a)\), the pure *de se* reading, and \((140b)\), one that’s also true in mistaken identity scenarios. The same should hold for \((138)\), i.e. in addition to the pure *de se* one given, we predict a reading that is true if I think, “that guy will win”, unaware that I’m pointing to myself. However, I suspect that, due to certain epistemic peculiarities of the first person, indicative, present tense nature of the main clause of \((138)\), this impure reading is highly unlikely, or even pragmatically out.\(^{88}\) As shown by von Stechow (1982:28), dream reports bring out the semantic binding of agreeing first persons much clearer.\(^{89}\)

\[(141)\] I dreamt I was Brigitte Bardot and I kissed me \([\text{Lakoff } 1972:639]\)
\[\triangleright\ \text{DREAM}^*_i \lambda x [\text{bardot}(x) \land \text{kiss}(x, i)]\]

I’ve represented the reading I get for \((141)\) as uttered by Lakoff. It involves a dream where the dreamself is, or at least strongly resembles, Brigitte Bardot and as such kisses Lakoff, who, in this particular dream is seen as a third person character, through the eyes of the dreamself.\(^{90}\)

---

\(^{88}\)The peculiar status of first person *de re*/*de se* reports, and their connection to Moore’s paradox, "It is raining but I don’t believe it", are discussed by Atlas (to appear).

\(^{89}\)Cf. also Heim (1994) and Percus & Sauerland (2003b) for more data and theory concerning *de re*/*de se* dream reports.

\(^{90}\)Though I’ve never had a dream where I see a third person version of myself this is attested by people who dream they see themselves as if in a movie, from an impersonal, objective camera perspective. It was quite hard to find testimony of dreams of the more complex type required for the Lakoff report, i.e. seeing a version of yourself through another person’s eyes, but I managed to get some confirmation at http://www.dreamviews.com/forum/viewtopic.php?p=234258.
As described by Heim (ms) and Kratzer (1998; 2006), binding under agreement is more widely applicable than just to attitude reports. As it turns out, the above account of the de re vs. pure de se ambiguity generalizes to an account of the strict and sloppy readings of pronouns in only complements (see 2.4.2.2 for our syntax/semantics of only):

(142) Only Ligia did her homework
      a. ⊢ ONLYX[x = 1][did_homework_of(x,x)]
      b. ⊢ ONLYX[x = 1][did_homework_of(x,1)]

In (142) the possessive pronoun’s surface features (3.sg.f) match those of the subordinating quantifier (Only Ligia, which derives those features from its restrictor, Ligia). By binding under agreement we may then remove the features of the pronoun in the quantifier’s scope and replace it, at lf, with an appropriately bound variable introduced by that quantifier. Incidentally, since we’ve already set up the our only as a traditional, variable binding quantifier, there’s no need for an extra λ here, and we can just pick x to replace the possessive. This gives us (142a), according to which Ligia is the only member of the set of students doing their own homework. If we forgo the binding option we get (142b), where Ligia is the only person who did Ligia’s homework.

Crucially, this same ambiguity applies to the first person variant:

(143) Only I did my homework [(Heim ms)]
      a. ⊢ ONLYX[x = i][did_homework_of(x,x)]
      b. ⊢ ONLYX[x = i][did_homework_of(x,i)]

The intuitive possibility of the sloppy reading in (143a), which is predicted by binding under agreement, supports the, at first preposterous sounding, idea (hastily rejected in 2.4.3.1, and e.g. by Partee (1989)\(^\text{91}\)) that even the paradigmatic indexical, an English first person pronoun, is sometimes interpreted as a bound variable rather than a directly referential device.

Now recall that the difference between Schlenker’s system and Heim’s was the fact that Schlenker maintains that embedded, shifted indexicals and pronominal elements like PRO and he under agreement, are semantically indexical in the sense that, even after feature deletion (if applicable), they get their reference from a context, viz. a quantified context variable, made

\(^{91}\)In fact, Partee claims only that a first person pronoun cannot be bound by a quantifier and gives the following example of a bound I in a relative clause:

(i) I’m the only one around here who will admit that I could be wrong[(Partee 1989)]
available by a monstrous attitude operator. Since there is no candidate for a monster to be found in examples like (143), it seems that indeed the deletion rules are independent of attitude reporting and context shifting. This has led von Stechow to the conclusion that we don’t really need monsters, we can let the deletion under agreement do all the work, i.e. turning indexicals into bound variables to derive pure de se readings.

2.4.3.5 A monster-free alternative?

Inspired by the Heim/Kratzer binding under agreement principle, Von Stechow (2001; 2002) sets out to defend Kaplan against Schlenker’s monsters. In his own words, he wants to show that:

while Kaplan (1989) is wrong in his claim that pronouns such as I or you are always directly referential, Schlenker’s (1999) criticism of Kaplan’s Prohibition against Monsters is not warranted. Schlenker’s attitudes are arguably purely intensional operators. There is no expression of natural language that denotes a genuine monster.

But if attitudes are not monsters how come they seem to sometimes shift the context of interpretation of an indexical? Short answer:

Attitudes [...] may bind the egocentric variables of their verbal complement under agreement and therefore may produce bound person/tense/mood readings. Under these circumstances the variables lose their referential character entirely.

The main idea, apparently, is still that of feature deletion under agreement, as discussed in the previous subsection: indexicals may be represented as mere bound variables when their surface features agree with those of the reporting verb.

Now, to rival Schlenker’s system, von Stechow integrates this principle in a more comprehensive semantics of attitude and speech reports that can handle all varieties of shifty and rigid indexicals. First of all, following Schlenker, he uses a many-sorted predicate logical extension with explicit quantification over individuals, worlds and times. Then, instead of context variables and monstrous quantifiers, he uses λ’s to represent ‘de se-binding’ in belief reports, following Heim’s (1994) analysis of (140), which in turn follows Chierchia’s (1989). As shown in 1.2.3.2 already, the properties self-ascribed in belief reports are not merely properties of individuals, but rather of contexts, construed as person-world-time triples. This corresponds also to
the observed fact that attitudinal de se binding can occur in exactly those 3 domains: individual, temporal, and modal. Where Schlenker quantifies over context variables, von Stechow heads his complements with a complex $\lambda x\lambda w\lambda t$ abstractor to get self-ascribable complements.

(144) Ellsworth believes she’s going to win

\[ a. \quad \text{BEL}^e_{e,n,t} \lambda x'\lambda w'\lambda t' [\exists t'' > t' [\text{win}(x',w',t'')] ] \]
\[ b. \quad \text{BEL}^e_{e,n,t} \lambda x'\lambda w'\lambda t' [\exists t'' > t' [\text{win}(e,w',t'')] ] \]

Semantically, we’re getting the same as before with Schlenker’s context quantifier, cf. (126):

\[
\left[ (144a) \right] = 1 \quad \text{iff} \quad \left[ (144a) \right] [x \mapsto a, w \mapsto c, t \mapsto c] = 1
\]
\[
\left[ \lambda x'\lambda w'\lambda t' [\exists t'' > t' [\text{win}(x',w',t'')] ] \right] [x \mapsto a, w \mapsto w', t \mapsto t] = \left[ \text{Bel}^c(\ell\text{llsworth}, w_c, t_c) \right] [x \mapsto a, w \mapsto w', t \mapsto t_c] = 1
\]
\[
\text{iff} \quad \text{for all belief alternatives } c' \in \text{Bel}^c(\ell\text{llsworth}, w_c, t_c), \text{ there is a } t > t_c \text{ at which } a_{c'} \text{ wins (in } w_{c'})
\]

In short, the logical representation and its semantics are but minimally different from Schlenker’s: abstracting $x$, $w$, and $t$ out of attitude embedded clauses comes out equivalent to having the attitude quantify over $c$’s (denoting the same $(a,w,c)$ triples), as long as we assume that certain referential elements may be bound by such a local parameter, while others are not, i.e. we need some version of Schlenker’s filtering mechanism, p. 160. Since both theories are in need of a deletion under agreement principle as well, what is the difference? As von Stechow himself observes, the differences are quite subtle, the crucial point is the way features are deleted, or rather, what’s left after a deletion. We’ve said it before, for Schlenker, the most straightforward grammaticalization of a de me report is Amharic $I$, that is to say, a first person form. Agreement principles are needed to ensure that third person logophors, pronouns and PRO can denote the belief context’s center as well.

For von Stechow on the other hand, pure de se readings correlate with a plain bound variable. This always requires feature deletion, for Amharic first as well as for English third person pronouns. Obviously then, something more than just feature agreement is needed, for the third person matrix subject John and the embedded first person in (93) do not agree in person. Von Stechow’s solution is to stipulate an extra, agreement independent deletion rule, called the Amharic Parameter (von Stechow (2002:25), ascribed to Heim (ms)), that allows Amharic attitude verbs to delete and bind embedded first
person features at will.

Thus, with an impressive array of morphosyntactic tools, von Stechow, like Schlenker, succeeds in accounting for the full range of data. As presented here, the differences between them are rather small, but to do them justice, a more thorough reconstruction and comparison of their approaches is necessary. For lack of space, we’ll have to leave it at this, however. The point on which I wish to improve on these approaches in the following is not empirical coverage—I’ll settle for equal coverage—but mostly methodological. The general motivation will be to shift the work from morphosyntax (agreement, deletion, movement, etc.) to pragmasemantics (dynamic interpretation, presupposition resolution, unification, etc.). The end result will be an account with a thoroughly transparent syntax, and a properly context-dependent interpretation module relying heavily on presupposition resolution.

2.A Many-sorted predicate logic

One can reduce the intensional predicate logic of 1.A to an extensional one by allowing quantification over possible worlds in the object language. The result is a two-sorted theory, indicating that there are two sets of points (individuals and worlds) in the model, and two corresponding types of variables and constants. Let’s go through a three sorted version with individuals, worlds and times.

2.A.1 Syntax

(145)  a. Primitive symbols

   (i) Variables: \( \text{Var} = \text{Var}^D \cup \text{Var}^W \cup \text{Var}^T \), i.e.:  
        over individuals: \( \text{Var}^D = \{x, x', y, \ldots\} \)  
        over worlds: \( \text{Var}^W = \{w, w', \ldots\} \)  
        over times: \( \text{Var}^T = \{t, t', \ldots\} \)  

   (ii) Constants: \( \text{Con} = \text{Con}^D \cup \text{Con}^W \cup \text{Con}^T \)  

   (iii) Terms: \( \text{Term}^\sim = \text{Var}^\sim \cup \text{Con}^\sim \)  

   (iv) For each \( k, l, m \) a set of \( k, l, m \)-place predicate symbols: \( \text{walk}, \text{loser} \) (1-1-1-place); \( \text{love}, = \) (2-1-1).

b. Atomic formulas

   (i) If \( \Pi \) is an \( k, l, m \)-place predicate and \( \tau_1, \ldots, \tau_k \in \text{Term}^D \), \( \tau_{k+1}, \ldots, \tau_{k+l} \in \text{Term}^W \), \( \tau_{k+l+1}, \ldots, \tau_{k+l+m} \in \text{Term}^T \), then \( \Pi(\tau_1, \ldots, \tau_{k+l+m}) \) is a formula.

c. Complex formulas
Chapter 2. Attitude reports

(i) If $\varphi$ and $\psi$ are formulas and $\xi \in \text{Var}$, then $\neg \varphi$, $[\varphi \land \psi]$, and $\exists \xi \varphi$ are formulas.

Free variables, $FV(\varphi) \subseteq \text{Var}$, defined as in 1.A.3.

2.A.2 Semantics

Model: $\mathfrak{M} = (D, W, T, \Pi)$

Assignment function: a partial function $f : \text{Var} \rightarrow D$, such that if $\xi \in \text{Var}^{D/W/T} \cap \text{Dom}(f)$, then $f(\xi) \in D/W/T$

And an assignment is proper for $\varphi$ iff $FV(\varphi) \subseteq \text{Dom}(f)$, as before.

Interpretation function: maps every $\gamma \in \text{Con}^{D/W/T}$ to a concept in $(D/W/T)^W$, and every $k,l,m$-predicate to a property in $(\varphi(D^k \times W^l \times T^m))^W$

Then, the Tarskian truth definition. In the following, $f$ as always denotes a proper assignment of the formula beneath it.

a. Primitive symbols:
   (i) If $\tau \in \text{Term}^{D/W/T}$, then $[\tau]^f \in D/W/T$
   (ii) If $\Pi$ is a $k$-$l$-$m$-place predicate symbol, $[\Pi]^f \in (\varphi(D^k \times W^l \times T^m))^W$

b. Atomic formulas:
   (i) $[\Pi(\tau_1, \ldots, \tau_{k+l+m})]^f = 1$ iff $\langle [\tau_1]^f, \ldots, [\tau_{k+l+m}]^f \rangle \in [\Pi]^f$ (otherwise = 0)

c. Complex formulas
   (i) $[\neg \varphi]^f = 1 - [\varphi]^f$
   (ii) $[\varphi \land \psi]^f = [\varphi]^f \cdot [\psi]^f$
   (iii) $[\exists \xi \varphi]^f = 1$ iff there is a $g \supseteq \{\xi\}$ s.t. $[\varphi]^g = 1$

2.A.3 Schlenker’s extensional Kaplan

Add a set of contexts $\mathcal{C}$ to the model and add a $c$ as extra evaluation parameter to each semantic value: $[\ldots]^f_c$. This context parameter is simply passed through until we hit an indexical.

Indexicals:

a. Syntax:
   (i) $i \in \text{Con}^D$ (I)
   (ii) $a \in \text{Con}^W$ (actually)
2.A Many-sorted predicate logic

(iii) \( n \in Con^T \) (now)

b. Semantics

(i) \([i]^f,c = a_c\), the speaker/agent of \( c \)
(ii) \([\mathbf{a}]^f,c = w_c\), the world of \( c \)
(iii) \([n]^f,c = t_c\), the time of \( c \)

Truth in a context is defined as truth in a context with respect to an assignment that maps the free variables \( w \) and \( t \) to \( w_c \) and \( t_c \) respectively. In other words, you can use these two special variables in your formula to denote, when free, the actual time and world, which is useful for emulating the index/evaluation parameters of the intensional system. We define \( f_c = f[w\mapsto w_c][t\mapsto t_c] \) (an assignment compatible with context \( c \)).

Now, \( \text{Bel}^{\text{int}} \) is an intensional operator, which, in this system comes down to a world quantifier:

\[
\text{Bel}^{\text{int}}[\text{Bel}^{\text{int}}(\tau,w,t)c\varphi]^f,c = 1 \text{ iff there is a character } C \text{ s.t. } \\
\delta C \supseteq \text{Bel}^{C}([\tau]^f,c,f(w),f(t)) \text{ and for all } w' \in W: \quad [\varphi]^{f[w'\mapsto w'],c} = C([\tau]^f,c,f(w),f(t))(w')
\]

2.A.4 Adding contexts to the language

(153) Context variables: \( \text{Var}^C = \{c, c', \ldots\} \)

(154) Functional indexicals: for every \( c \in \text{Var}^C \):

a. Syntax:

(i) \( a_c \in \text{Term}^D(1) \)
(ii) \( w_c \in \text{Term}^W \) (actually)
(iii) \( t_c \in \text{Con}^T \) (now)

b. Semantics

(i) \([a_c]^f = a_{f(c)}\), the speaker/agent of \( f(c) \)
(ii) \([w_c]^f = w_{f(c)}\), the world of \( f(c) \)
(iii) \([t_c]^f = t_{f(c)}\), the time of \( f(c) \)

Note that the definition of truth does not need any relativization to context or index parameters, this is now taken over by the assignment function. Truth in a context \( c \) is therefore defined as truth with respect to a \( c \)-compatible assignment \( f_c = f[c\mapsto c][w\mapsto w_c][t\mapsto t_c] \).

Belief reports quantify over contexts:

\[
\text{Bel}^{\text{int}}[\text{Bel}^{\text{int}}(\tau,w,t)c'\varphi]^f,c = 1 \text{ iff for all } c' \in \text{Bel}^{C}([\tau]^f,c,f(w),f(t)): \quad [\varphi]^{f[c'\mapsto c']} = 1
\]
2.B Typed $\lambda$-calculus

In type theory, every expression has a type which determines whether that expression refers to an individual, a predicate, a relation or something higher-order. The types therefore determine whether or not two or more expressions together can form a new one. Furthermore, at a semantic level all such syntactic combinations correspond to functional application.

2.B.1 Types

The set of types for standard extensional type theory is defined as follows:

(156) a. $e$ is a type (the type for referring to entities)
     b. $t$ is a type (the type for referring to truth values)
     c. if $\tau$ and $\sigma$ are types then so is $\langle \tau, \sigma \rangle$

A notational convention: association of type brackets is to the right, e.g.:

(157) $\langle e, \langle \langle t, t \rangle, t \rangle \rangle = e(tt)t$

2.B.2 Syntax

(158) The set of well-formed expressions, $\text{Exp}$:

a. Basic expressions (of a certain type):
   (i) $\text{Con}_\tau$ is the (possibly empty) set of constants of type $\tau$
   (ii) $\text{Var}_\tau$ is the (infinite) set of variables of type $\tau$
   (iii) $\text{Exp}_\tau \supset \text{Con}_\tau \cup \text{Var}_\tau$

b. Complex expressions:
   (i) If $\alpha, \beta \in \text{Exp}_\tau$ and $\xi \in \text{Var}_\tau$, then $\lnot \alpha, [\alpha \land \beta], \exists \xi \alpha \in \text{Exp}_\tau$
   (ii) If $\alpha, \beta \in \text{Exp}_\tau$, then $[\alpha = \beta] \in \text{Exp}_\tau$
   (iii) If $\alpha \in \text{Exp}_\tau$, $\beta \in \text{Exp}_{(\tau, \sigma)}$, then $[\beta(\alpha)] \in \text{Exp}_\sigma$
   (iv) If $\alpha \in \text{Exp}_\sigma$, $\xi \in \text{Var}_\tau$, then $\lambda \xi \alpha \in \text{Exp}_{(\tau, \sigma)}$

(159) $\text{Con} = \bigcup_\tau \text{Con}_\tau, \text{Var} = \bigcup_\tau \text{Var}_\tau, \text{Exp} = \bigcup_\tau \text{Exp}_\tau$

There’s couple of conventions that make our formulas easier to read: (i) we abbreviate statements like ‘$\alpha \in \text{Con}_\tau$’ by subscripting the first occurrence of a basic expression in a formula with the type to which it belongs, (ii) superfluous brackets are omitted, and (iii) the higher the type, the fancier the lettering: $x, X, X \ldots$. To further the intended denotation of constants and variables we always try to choose appropriate letters and even abbreviated words and phrases. For example:
2.B Typed $\lambda$-calculus

Further, (iv) consecutive arguments may be grouped together:

$$[[\text{love}_{ext}(x_e, y_e)] = [[\text{love}_{ext}(x_e)](y_e)]$$

2.B.3 Models

A model must contain denotations for every type of expression. Given only a basic domain of individuals $D$ and a set of truth values, a model determines an entire domain function mapping each type $\tau$ onto a domain of individuals $D_{\tau}$ suitable for that type:

$$\begin{align*}
D_e &= D \\
D_t &= \{0, 1\} \\
D_{\langle \tau, \sigma \rangle} &= D_{\sigma}^{D_{\tau}}
\end{align*}$$

$$D = \bigcup_{\tau} D_{\tau}$$

A model must also contain an interpretation function to interpret the constants of the language:

$$\mathcal{I} : \text{Con} \to D \text{ s.t. for all } \alpha \in \text{Con}_{\tau} : \mathcal{I}(\alpha) \in D_{\tau}$$

$$\text{A model } \mathfrak{M} \text{ is a pair } \langle D, \mathcal{I} \rangle$$

2.B.4 Semantics

We define the set of free variables of a formula recursively:

$$\begin{align*}
\text{a. If } \xi \in \text{Var}, \text{ then } FV(\xi) &= \{\xi\} \\
\text{b. } FV(\alpha \land \beta) &= FV(\alpha = \beta) = FV(\alpha(\beta)) = FV(\alpha) \cup FV(\beta) \\
\text{c. } FV(\neg \alpha) &= FV(\alpha) \\
\text{d. } FV(\exists \xi \alpha) &= FV(\lambda \xi \alpha) = FV(\alpha) \setminus \{\xi\}
\end{align*}$$

Semantic values of arbitrary expressions are given relative to an assignment function:

$$\text{A partial function } f : \subseteq \text{Var} \to D \text{ with for each } \xi \in \text{Var}_{\tau}, f(\xi) \in D_{\tau} \text{ is called a proper assignment for an expression } \alpha \text{ iff } FV(\alpha) \subseteq \text{Dom}(f)$$

Interpretation is defined as follows (it is assumed that $f$ is a proper assignment for the formula it is superscripting):

$$\begin{align*}
\text{a. Basic expressions:}
\end{align*}$$
(i) If \( \alpha \in \text{Con} \), then \( \llbracket \alpha \rrbracket^f = \mathbb{I}(\alpha) \)
(ii) If \( \xi \in \text{Var} \), then \( \llbracket \xi \rrbracket^f = f(\xi) \in D \)

b. Complex expressions:
(i) \( \llbracket \neg \alpha \rrbracket^f = 1 - \llbracket \alpha \rrbracket^f \)
(ii) \( \llbracket \alpha \land \beta \rrbracket^f = \llbracket \alpha \rrbracket^f \cdot \llbracket \beta \rrbracket^f \)
(iii) \( \llbracket \exists \xi \alpha \rrbracket^f = 1 \text{ iff there is a } g \supseteq \xi \text{ f s.t. } [\alpha]^g = 1 \)
(iv) \( \llbracket \alpha = \beta \rrbracket^f = 1 \text{ iff } \llbracket \alpha \rrbracket^f = \llbracket \beta \rrbracket^f \)
(v) \( \llbracket \beta(\alpha) \rrbracket^f = \llbracket \beta \rrbracket^f \llbracket \alpha \rrbracket^f \)
(vi) \( \llbracket \lambda \xi \tau \alpha \rrbracket^f : D_\tau \rightarrow D_\sigma \) and for all \( d \in D_\tau \), \( \llbracket \lambda \xi \alpha \rrbracket^f (d) = \llbracket \alpha \rrbracket^f[\xi \mapsto d] \)

2.B.5 Conversions

The semantics validates three well-known syntactic conversion operations, but first a notational shorthand:

(169) \( \alpha_\sigma[\xi_\tau \mapsto \beta_\tau] \) is the expression that results from replacing in \( \alpha \) all free occurrences of the variable \( \xi \) by the expression \( \beta \).

The annoying thing is that if we substitute a variable in \( \alpha \) for a complex expression with free variables of its own, these free variables may become bound by quantifiers or lambdas already present in \( \alpha \). The ‘ifs’ below are meant to exclude this possibility.

(170) Renaming bound variables/\( \alpha \)-conversion:
\( \lambda \xi_\tau \alpha_\sigma = \lambda \nu_\tau \alpha_\sigma[\xi_\tau \mapsto \nu_\tau] \), if \( \text{FV}(\lambda \xi \alpha) = \text{FV}(\lambda \nu \alpha[\xi \mapsto \nu]) \) (idem for \( \exists \) or \( \forall \) instead of \( \lambda \))

(171) Functional application/\( \beta \)-reduction:
\( \lambda \xi_\tau \alpha_\sigma(\beta_\tau) = \alpha_\sigma[\xi_\tau \mapsto \beta_\tau] \), if \( \text{FV}(\lambda \xi \alpha(\beta)) = \text{FV}(\alpha[\xi \mapsto \beta]) \)

(172) Extensionality/\( \eta \)-reduction:
\( \lambda \xi_\tau(\alpha_{(\tau, \sigma)})(\xi_\tau) = \alpha \)

2.B.6 Product types

(173) Types:
If \( \sigma_1 \ldots \sigma_n \) are types, then so is \( \sigma_1 \times \ldots \times \sigma_n \)

(174) Syntax:
a. If \( \alpha_1 \in \text{Exp}_{\sigma_1}, \ldots, \alpha_n \in \text{Exp}_{\sigma_n}, \text{ then } \langle \alpha_1, \ldots, \alpha_n \rangle \in \text{Exp}_{\sigma_1 \times \ldots \times \sigma_n} \)
b. If \( \alpha \in \text{Exp}_{\sigma_1 \times \ldots \times \sigma_n}, \text{ then } (\alpha)_1 \in \text{Exp}_{\sigma_1}, \ldots, (\alpha)_n \in \text{Exp}_{\sigma_n} \)
(175) Structured propositions:

If \( \alpha_1, \ldots, \alpha_n, \beta \in \text{Exp} \) s.t. \( \beta(\alpha_1, \ldots, \alpha_n) \in \text{Exp}_t \), then \( \langle \langle \alpha_1, \ldots, \alpha_n \rangle, \beta \rangle \) is a structured proposition

(176) Free variables:

a. \( FV(\langle \alpha_1, \ldots, \alpha_n \rangle) = \bigcup_n FV(\alpha_n) \)

b. \( FV((\alpha)_i) = FV(\alpha) \)

(177) Domains:

\( D_{\sigma_1 \times \ldots \times \sigma_n} = D_{\sigma_1} \times \ldots \times D_{\sigma_n} \)

(178) Semantics:

a. \( \llbracket \langle \alpha_1, \ldots, \alpha_n \rangle \rrbracket^f = \langle \llbracket \alpha_1 \rrbracket^f, \ldots, \llbracket \alpha_n \rrbracket^f \rangle \)

b. \( \llbracket (\alpha_{\sigma_1 \times \ldots \times \sigma_n})_i \rrbracket^f = (\llbracket \alpha \rrbracket^f)_i \) (i.e., the \( i \)th coordinate of the \( n \)-tuple denoted by \( \alpha \))

2.B.7 Intensional types and belief

(179) If \( \tau \) is a type, then so is \( \langle s, \tau \rangle \)

We assume that sentence representation have intensional types, \( \langle s, t \rangle \), so we define conjunction and other sentential operators only for these. Lambda abstraction remains the same as before. \( \lor \) is the operator that makes an intensional term ‘less intensional’.

(180) a. If \( \alpha \in \text{Exp}_e, \xi \in \text{Var}_\tau, \varphi, \psi \in \text{Exp}_{\langle s,t \rangle} \), then \( \neg \varphi, \varphi \land \psi, \exists \xi \varphi, \Box \varphi \in \text{Exp}_{\langle s,t \rangle} \)

b. If \( \alpha \in \text{Exp}_{\langle s,t \rangle} \) then \( \lor \alpha \in \text{Exp}_\tau \)

(181) \( \mathcal{M} = \langle D, W, \llbracket \rangle \rangle \)

(182) \( D_s = W \)

(183) a. \( \llbracket \neg \varphi \rrbracket^f = W \setminus \llbracket \varphi \rrbracket^f \)

b. \( \llbracket \varphi \land \psi \rrbracket^f = \llbracket \varphi \rrbracket^f \cap \llbracket \psi \rrbracket^f \)

c. \( \llbracket \exists \xi \varphi \rrbracket^f_w = 1 \) iff there is a \( g \supseteq \xi f \) with \( \llbracket \varphi \rrbracket^g_w = 1 \)

d. \( \llbracket \Box \varphi \rrbracket^f_w = 1 \) for all \( w'Rw: \llbracket \varphi \rrbracket^f_{w'} = 1 \)

(184) \( \llbracket \forall \alpha \rrbracket^f_w = \llbracket \alpha \rrbracket^f_w (w) \)

Belief as property self-ascription:\(^{93}\)

[^93]: Chapter 2 explores a host of different variations on \( \text{BEL}^* (\text{BEL}, \text{BEL}^{e \times e}, \text{BEL}) \) which we will not repeat here.
If $\beta \in \text{Exp}_{\langle e, d \rangle}$ then $\llbracket \text{Bel}_\alpha \beta \rrbracket^f (w) = 1$ iff all $\langle w, d \rangle$ with $\llbracket \beta \rrbracket^f_w (d) = 1$ are in $\text{Bel}(\llbracket \alpha \rrbracket^f_w, w)$ (iff $\alpha$ self-ascribes the property $\beta$)
Chapter 3

A semantics of attitude reports

Now it’s time to present an alternative theory, a formal framework that ultimately maps belief reporting sentences onto their interpretations. As mentioned before, we will pick up the relational attitude analysis in which first person *de se* and *de re* are just two examples of acquaintance relations giving rise to a relational attitude. Note that we have also considered the fully compositional semantics provided by Kaplan (1989). That account and a semantics based on the relational view of belief share the unified treatment of *de re/de se* and the idea that natural language reports are underspecified for the exact subtype of the underlying belief. Unfortunately, Kaplan’s analysis, originally intended for speech reports in indirect discourse, turned out to be fundamentally flawed as a semantics for natural language belief reporting.

Moreover, Kaplan’s analysis was shown to be incompatible with what I consider a *sine qua non* for the semantics of belief reporting: taking context dependence into account. Semantic history, and in particular the tiny fraction described in 2.3.2, has shown that a formal semantics of attitude reports may benefit a lot from taking into account the context dependence of acquaintance relations. However, despite some promising first steps, the classical report semantics in terms of relational attitudes with contextualized acquaintance suffers a number of serious defects, as pointed out in subsequent sections of chapter 2. The objections in a nutshell: (i) not even the relational semantics itself can be cashed out in a fully compositional fashion, let alone any contextualized version; (ii) with or without the contextualization of acquaintance, the relational semantics fails to make the right predictions with respect to reports embedded under quantifiers; (iii) in fact, the underlying assumption that the various types of *de re* and *de se* belief can be taken together and given a unified/underspecified semantics has become threatened by more recent data showing that natural languages can select for specific pure *de se* types of attitude; (iv) occurrences of shifted indexicality pose
problems for relational (and Kaplanian) analyses, which in effect evaluate all attitude embedded referential terms as if they had wide scope over the attitude (Principle 2).

In the following I aim to solve these problems, developing an enhanced version of the relational account with context dependent acquaintance. However, to make it outperform Schlenker’s and von Stechow’s state of the art rival accounts, a rather drastic change is needed. Over the next sections I propose a change of framework, from static to dynamic semantics, Discourse Representation Theory to be precise, to fully exploit context dependence.

### 3.1 Discourse Representation Theory

#### 3.1.1 Varieties of dynamic semantics

Increasing awareness of the importance of context dependence for the study of natural language semantics, and the long standing problems and puzzles posed by one of its main manifestations (presupposition) in particular,\(^1\) have inspired the rise of a new paradigm in semantics, called *dynamic semantics*. The first fully developed version was Kamp’s (1981) *Discourse Representation Theory* (DRT), closely followed by the independent *File Change Semantics* (FCS) of Heim (1982; 1983a), and later Groenendijk & Stokhof’s (1991) *Dynamic Predicate Logic* (DPL). More variations, extensions and improvements on these exist, but the defining characteristic of the new paradigm is that *context change* replaces truth and reference as the central semantic notions. Thus, classical truth conditional semantics is now often referred to as *static semantics*. This is not to say that classical semantics cannot deal with context dependence or context change—as witness for example Kaplan’s (1989) work on deictic context dependence and Stalnaker’s (1978) work on context change in an arguably static framework\(^2\)—nor that dynamic semantics cannot account for the fact that utterances express propositions and are (often) true or false, as will be shown below.

In 3.1.2 I make a case for dynamic semantics, and for DRT in particular. Then, in 3.1.3, I give a detailed, formal presentation of the logic of standard DRT. After the basic framework is laid out, I’ll proceed to add the necessary features needed to accommodate my version of the relational

---

\(^1\)Other manifestations of context dependence are still a mess (implicature), or already quite satisfactorily accounted for in static semantics (indexicality).

\(^2\)In Stalnaker’s seminal writings from the seventies, utterances are taken to express static propositions, or rather two-dimensionally generalized versions thereof called *propositional concepts*, which are then used to narrow down a context set representing the growing common ground (by intersecting with it).
attitude semantics. First, presuppositions (3.2), and attitude operators with classical possible worlds semantics 3.2.3. Then some more advanced concepts in 3.3, like layers to separate rigid and descriptive content (with matching two-dimensional semantics), and the addition of higher-order expressions and unification thereof. In the resulting DRT extension we then implement the contextualized relational attitude semantics.

3.1.2 Why DRT?

We shall be concerned exclusively with DRT, so let’s briefly consider what could be the reasons for adopting that framework, as opposed to the familiar, classical frameworks and the dynamic alternatives.

3.1.2.1 DRT vs. static semantics

As implied by its name, DRT is a theory of discourse interpretation, which means that we shift the focus from isolated sentences to sentences uttered in the context of a discourse. One of the original motivations was to account for the interpretation of intersentential anaphora. Consider the following mini discourse:

(1) Ligia has a dog. Her husband abuses it.

To interpret the second sentence, we need the context provided by the first, otherwise, what would her refer to, or if?

As it turns out, many expressions in natural language exhibit this type of anaphoric context dependence, and classical logic is not very well suited to deal with it. In predicate logic we can easily represent the truth conditions of the first sentence, and of the whole discourse, but what exactly is the contribution of the second?

How to model the transition of one closed formula to the next? And should we map entire texts onto gigantic, monolithic formulas?

\[ \exists x \left[ \text{dog}(x) \land \text{own}(1, x) \right] + ???? \]

\[ \leadsto \exists x \exists y \left[ \text{dog}(x) \land \text{own}(1, x) \land \text{husband}(y, 1) \land \text{abuse}(y, x) \right] \]

To be fair, there are some ways to deal with intersentential anaphora in classical semantics. One is to represent anaphoric expressions as free variables, which means they are seen as referential expressions that get their interpretation from an assignment function. It is then further assumed that assignment functions are somehow fixed by an independent, unspecified pragmatic module that takes the context into account. We’ve already encoun-
tered an application of this maneuver in 2.3.2 where we also pointed out that it merely moves the real issue out of sight, into a pragmatic wastebasket. Besides, though it may work fine for (1)’s *her*, picking out Ligia, it’s already harder (though not impossible) to see how *it* would pick out the dog introduced by the existentially quantified NP inside the first sentence. Apparently, a solution along these lines would have to meddle with the semantics of indefinites as well, for apparently they are able to extend assignment functions beyond sentence boundaries. To put it differently, indefinites can pragmasemantically bind variables outside their syntactic scope. This is actually getting us quite close to the role of indefinites in dynamic frameworks, though the next example will show that still more is needed.\[^3\]

A much harder version of the above problem similarly revolving around anaphora to indefinite antecedents, concerns intrasentential but cross-clausal anaphora. This issue has become known as the problem of *donkey anaphora*, after its paradigmatic examples involving all manner of donkey molestation:\[^4\]

\[\text{(2) If a farmer owns a donkey, he beats it}\]

Again, it’s quite easy to write down adequate truth conditions for the whole, and for the first clause (*a farmer owns a donkey*), but now it’s even more of a mystery than for the intersentential case how conjoining the existential first and the anaphoric second clause can get us the overall universal reading:

\[
\begin{align*}
\vdash & \exists x \exists y [\text{farmer}(x) \land \text{donkey}(y) \land \text{own}(x, y)] + \text{???} \\
\sim & \forall x \forall y [\text{farmer}(x) \land \text{donkey}(y) \land \text{own}(x, y)] \rightarrow \text{beat}(x, y)
\end{align*}
\]

Presented like this it’s really a compositionality problem: it seems that the whole sentence’s LF is not compositionally built up from the LFs of its constituent clauses and an implication. DRT tries to fix this by offering a different syntax and semantics for (existential) quantification and implication (in addition to a new way of representing and binding anaphors).

\[^3\]The second classical solution to the intersentential anaphora problem is Evans’s (1977) E-type analysis, according to which anaphoric expressions are to be represented as full definite descriptions whose content is somehow copied from the context. In paraphrase, (1) is analyzed as *There exists a dog that Ligia has. Ligia’s husband abuses the dog that Ligia has*. In this simple form it has the same defects as the free variable approach: (i) we need a pragmatic black box that picks an appropriate description from the context, and (ii) donkey sentences remain problematic (though Heim (1990) introduces some nontrivial extensions to make this work).

\[^4\]The original examples (with donkeys) originated in the Middle Ages. The problem was reintroduced into the philosophy of language by Geach (1962).
3.1 Discourse Representation Theory

To repeat, in dynamic semantics we take a different view of meaning and interpretation: instead of (a formula representing) a set of worlds, each clause contributes a context change potential, specifying how it would affect the context it is uttered in. To see how the idea of context change is cashed out in DRT, we must first take a look at the notion of context. In DRT, contexts are represented by a special kind of formulas called Discourse Representation Structures (DRS), to be defined in 3.1.3. As we will see, DRSs can be translated back into predicate logic, or given their own static interpretation directly, but their syntax is designed in such a way that we can add the context changing information conveyed by an utterance in a relatively straightforward, algorithmic fashion. The result of this process, which comprises the monotonic information growth and the resolution of anaphora, among other things, will be a new context DRS, to which we can add the next sentence and so on for the rest of the discourse.

To make this work smoothly, van der Sandt & Geurts (1991) propose that the DRS syntax used to represent contexts, be used to model context change potentials as well. This requires some adaptation of the DRS language; we need a way to preliminarily represent anaphors and other expressions that derive their meaning from the discourse context. To distinguish them, the structures representing context change potentials are called preliminary DRSs. The dynamic interpretation of an utterance in a discourse is thus cashed out as a merge operation of a context DRS with the current sentence’s preliminary DRS after which anaphora are resolved. In 3.1.3 and 3.2 we will see in some detail, and with examples, how this works exactly.

3.1.2.2 DRT vs. the dynamic alternatives

Now that we have outlined the idea behind DRT’s implementation of the dynamic aspect of discourse meaning, it might be interesting to consider briefly what sets apart DRT from the other types of dynamic semantics, before going into the details of formalization. The most heavily debated feature peculiar to DRT is its reliance on representations. The whole interpretation process is formalized at this representational, syntactic level. It’s like an algorithm that manipulates formulas, taking a context DRS and a preliminary DRS as input and giving a new DRS as output. In the previous chapter we have similarly used the language of predicate logic as a convenient way to talk about truth conditions. In both cases the formal language serves as an intermediary between natural language and semantic interpretation, though in the original DRT framework truly semantic interpretations are defined only for contexts. After the idea of stand-alone preliminary DRSs had caught on, they too
were quickly endowed with model theoretic interpretations. Unfortunately, the challenge of giving a truly semantic interpretation of preliminary DRSs that include adequate representations of presuppositions has not been met.

So in the case of preliminary DRS construction and transformation we see DRT operating at a syntactic level with no (obvious) counterpart in model theoretic semantics. Some philosophers and semanticists might take issue with this kind of thing, but it is fully compatible with the mentalistic philosophy behind DRT, which was really designed to model what goes on in the mind of a human (or computer) language user interpreting a discourse, at a suitably abstract level, obviously. In other words, familiar from 1.2.4.4, DRT primarily aims to capture the narrow aspect of meaning and interpretation, i.e. meaning as thought of by many psychologists and cognitive scientists. The only thing that stands between DRT’s representationalism and full-blown mentalistic sententialism (denounced in 1.1.2.3) is that DRT comes with a ‘wide’, Tarskian truth definition for its representations—though only for context representations that constitute the outputs of the procedural interpretation process. In short, given its roots in cognitive science, it’s not surprising that DRT essentially relies on representational data structures and their algorithmic manipulation. In dealing with presupposition resolution, more specifically with presupposition accommodation, we’ll again rely on symbol manipulation at the DRS level.

The DRT conception of meaning and representation is in stark contrast with the use of lf representations in the Montagovian tradition. There, the logical language is merely a matter of convenience, in the sense that it can be eliminated. This is because of the strong compositionality built into Montagovian systems (1.1.1.3): each syntactic rule corresponds to a semantic rule, so in giving the general rules of the system and a sentence’s syntac-

---

5 E.g. in terms of ccp s, by analogy with Heim’s semantic interpretation of FCS ‘file cards’ as context change potentials, modeled as operators on information states, i.e. sets of \( \langle \text{world, assignment} \rangle \)-pairs.

6 DRT’s founding father expresses his own defeatist attitude towards this challenge thus:

Unfortunately the possibility of such a semantics [for preliminary DRSs] is limited. […] [F]or those preliminary representations […] in which all presuppositions appear in the highest possible position […] an intuitively plausible model theoretic semantics can be stated without too much difficulty. But for representations with presuppositions in subordinate positions I haven’t yet been able to find a satisfactory semantics and I very much doubt that one is to be had. ([Kamp 2001:44])

7 Well, a specific type of cognitive scientist at least, perhaps to be identified with those who believe in symbolic or ‘good old-fashioned AI’.
tic tree we have already fixed the whole interpretation; nothing can really be gained by introducing an intermediate level of IIs in a logical language. Montague (1970a) actually demonstrated this by providing a fragment of English in which each syntactic constituent gets a model theoretic interpretation directly. Obviously this feat requires heavy semantic assumptions to intrude on the syntax (thereby clashing with the generative/Chomskyan tradition which preaches the autonomy of syntax in the determination of constituent structure), and a lot of work would have to be done to turn the theories of chapter 2 into genuinely non-representational ones in this strong sense. It is not even very likely that this can be done, so in effect many theories of natural language rely on some form of indispensable intermediate representation. For instance in the neo-Chomsky-Montagovian semantic tradition—Chierchia, von Stechow, Heim, Kratzer, and Schlenker, to name a few key theorists already encountered in the previous chapter—this level is the LF which is typically not compositionally derived from the surface form, but which does translate immediately to an If and thus to truth conditions. In the simplified toy versions of these theories sketched in chapter 2 we condensed If and LF into an appropriate extension of predicate logic, but I’m not convinced that that representational level can be dispensed with in favor of a rigorous syntactic analysis of surface structure à la Montague.

Now, back to dynamic semantics, where DRT competes with Heim’s FCS and Groenendijk & Stokhof’s DPL. Heim (1983a) introduces a representational level called file cards which are remarkably similar to DRSs in both form and function. So, in their representation of context, DRT and FCS do not really differ, especially since file cards, like DRSs, can be embedded into a genuine, static model. Moreover, preliminary DRSs also have an exact analogue in FCS, viz. logical forms (henceforth with capitalized initials, since the notion is closer to our LF than to If). However, the semantic philosophy behind FCS appears to be quite different from DRT’s, which comes to the fore when Heim starts filling in details and actual applications of her theory. From an early point on, file cards are identified with their model theoretic satisfaction sets. LF’s are similarly interpreted as functions from satisfaction sets to satisfaction sets (context change potentials), rather than from files to files (file change potentials). It’s obvious that Heim tries to bypass the representational level in favor of model theoretic entities. The Montagovian goal of

---

8It should be noted that satisfaction sets are sets of assignment functions, or something interdefinable with that, and thus not as purely non-representational as, say, propositions or even characters in static predicate semantics, for an assignment function crucially contains variables, which are symbolic entities. This representational aspect is retained in DPL which claims explicitly that it’s main advantage over DRT lies in its elimination of the representational level. This point was made by Geurts (1999).
eliminating representations is made explicit by Groenendijk & Stokhof (1991) whose DPL looks like normal predicate logic, syntactically, but with a dynamic model theoretic interpretation. The details of this system really don’t concern us here, the important thing is that it recycles the familiar syntax of classical predicate logic. The next logical step would be to see if this dynamicization of predicate logic extends to full higher order, intensional type theory, so that Montague’s fully compositional, non-representational system could be straightforwardly enhanced with a dynamic rather than static interpretation via a similarly eliminable type logic in the middle. Indeed, such a dynamic Montague Grammar was constructed by Groenendijk & Stokhof (1990).

It should be noted that the powerful presupposition theory that van der Sandt (1992) proposed for DRT requires movements at the representational level, which makes it unsuitable for the semanticized FCS and for DPL/DMG. Alternative presupposition theories have been developed for both, but their empirical predictions are quite distinct (see Beaver (2001) for an overview and defense of the non-representational view of presuppositions and Geurts (1999) for the opposing view, which is to be presented in 3.2). Tentatively adding presupposition theory as a plus of DRT, let’s sum up the rest of the advantages. Firstly, DRT has shown its merits in the unparalleled empirical coverage with its analysis of tense and aspect (Kamp & Reyle 1993). Then, with considerable extensions, (S)DRT is considered the state of the art in analyzing discourse relations (Asher & Lascarides 2003). Thirdly, there have been very promising attempts at solving the classic omniscience puzzles by using DRSs as models of the mental states of individuals. We will say something about this in 3.2.3, but note that things have gotten rather complicated here in a way that I cannot do justice to. In this area, the mentalism of DRT has come to play a major role, but for reasons of compatibility with the classic theories of attitude report semantics (e.g. the ones discussed in the previous chapter), my own proposal will not follow in this tradition, in line with my conviction that the studies of attitudes and of attitude reports should be separated as rigorously as possible (as stated in 1.1.2.3). I view DRSs as representing the common ground, rather than the mental state of the hearer. A final, more general advantage of the AI-inspired representationalism of DRT and its add-ons, like van der Sandtian presupposition theory, is the possibility of straightforward computational implementation, but we will not go in to that here (Blackburn & Bos 1999).

To summarize, DRT was originally based on a mentalistic conception of meaning, so interpretation is primarily cashed out in terms of symbolic operations encoding context change. The whole process is nonetheless restricted by the external world in the sense that the output after a clause’s
processing is subject to a static truth definition. Presuppositions are one area where the representational level can make a big difference. DRT’s rival FCS has similar data structures but puts the focus on their model theoretic semantic interpretations, thus forgoing the powerful representational theory of presuppositions that was developed for DRT. This Montagovian tendency of bypassing the symbolic level is taken up in DPL which forms the basis of a non-representational, fully compositional yet dynamic implementation of Montague’s program.

3.1.3 The language of DRT

In this we take a look at the details of the DRT formalism, i.e. the syntax and (static) semantics of the DRS language. I will closely follow all standard definitions and terminology, mostly from Kamp & Reyle (1993), including the two-dimensional box notation. In this subsection I concentrate on DRSs as a language for representing the common ground. Section 3.2 deals with natural language interpretation and common ground incrementation via preliminary DRSs.

3.1.3.1 Syntax

A DRS is an ordered pair consisting of a set of discourse referents (e.g. \(x, y\)), called the universe of the DRS, and a set of DRS conditions. Discourse referents in DRT serve the same function as variables in predicate logic, and the ones that occur in a DRS universe are bound within that DRS. Conditions come in two types, atomic, like atomic formulas, (e.g. \(\text{beat}(x, y)\)); and complex, i.e. conditions that themselves contain full DRSs (e.g. \(\neg \psi\)). The following notations go with these notions: \(\varphi = \langle U(\varphi), C(\varphi) \rangle \in DRS = DRef \times DCond\). In real examples I opt for the familiar notation which depicts DRSs as two-compartment boxes; universe on top, conditions below:

(3) Colby is in love with a philosopher

\[
\begin{array}{c|c}
\text{x} & \text{y} \\
\hline
\text{colby}(x) & \text{philosopher}(y) \\
\text{love}(x, y) \\
\end{array}
\]

Think of this as follows: the universe specifies that the discourse is about two individuals, and the three conditions further specify who these individuals are and what they do. An equivalent representation in predicate logic would look like this:
Chapter 3. A semantics of attitude reports

(4) \[ \exists x \exists y [\text{colby}(x) \land \text{philosopher}(y) \land \text{love}(x,y)] \]

The only difference with the predicate logical if we’d normally assign to the sentence in (3), is the proper name, which is not represented by a constant, but by a variable whose reference is restricted by a predicate. This is only a superficial difference though; with our implicit assumptions on proper name reference (uniqueness and rigidity, cf. chapter 1: 1.2.2.1 and footnotes 25 and 26) we can easily interdefine these: (i) \( c \sim x. \text{colby}(x) \); (ii) \( \text{colby}(x) \sim x = c \). I’ve simply chosen the traditional way of doing things in DRT, conforming to the intuitive idea that every entity that figures in the discourse is made available for future anaphora by introducing a corresponding discourse referent (or by picking up an already existing one). There’s more to be said about proper names, but only after we’ve introduced presupposition resolution in 3.2.

Simply put, the interpretation of new sentences in a discourse proceeds by adding new discourse referents to the existing DRS universe and/or adding new conditions to its condition set. How this is done exactly is discussed 3.2, but the end result of adding, say \textit{He gave her a present} is the following incrementation:

(5) Colby is in love with a philosopher. He gave her a present.

\[
\begin{array}{c|c|c}
   & x & y \\
\hline
\text{colby}(x) & \text{philosopher}(y) & \text{love}(x,y) \\
\hline
\end{array}
\]

\[
\begin{array}{c|c|c}
   x & y & z \\
\hline
\text{colby}(x) & \text{philosopher}(y) & \text{love}(x,y) \\
\text{present}(z) & \text{give}(x,z,y) \\
\hline
\end{array}
\]

So much for atomic conditions. Negation, implication and disjunction are represented as operators on DRSs:

(6) If \( \varphi \) and \( \psi \) are DRSs, then \( \neg \varphi, \varphi \Rightarrow \psi \) and \( \varphi \lor \psi \) are DRS conditions

Note that we don’t need a conjunction, because it’s already given in our box notation, where a sequence of conditions is going to mean what predicate logic would express with conjunction, as shown by the example translations into predicate logic above. The same holds for existential quantification: putting a discourse referent in the top compartment of a DRS is the DRT way of existentially quantifying over it. We should however not hastily conclude the work is done if we’ve defined just a negation to go with our implicit conjunction and existential quantifier, because, as we will see in the
well-formedness definition and the semantics, DRT will bestow some extra, dynamic features on our implication which might invalidate some definitively useful equivalences of propositional logic (e.g. in DRT \( \neg\neg\varphi \) is statically but not dynamically equivalent to \( \varphi \), in a sense to be made precise below).

An example:

(7) Ligia doesn’t know Colby, or, if she does know him, she doesn’t like him

\[
\begin{array}{c}
\text{x y} \\
\text{ligia(x)} \\
\text{colby(y)} \\
\end{array}
\]

\[
\begin{array}{c}
\lor \\
\neg\text{know}(x,y) \\
\text{know}(x,y) \\
\text{like}(x,y) \\
\end{array}
\]

Let’s use this rather ‘boxy’ DRS to demonstrate a handful of formal notions useful for describing the hierarchical nesting structure of DRSs. First, we say that the biggest box is the main or global DRS, the others are embedded within it. More precisely, there is a relation \( \sqsubset \), pronounced “is immediately subordinate to” on DRSs,\(^{10}\) according to which, for instance:

(8)\[
\begin{array}{c}
\text{like}(x,y) \\
\neg\text{like}(x,y) \\
\text{know}(x,y) \\
\text{like}(x,y) \\
\end{array}
\]

For future applications the so-called transitive reflexive closure of \( \sqsubset \), denoted \( \sqsubseteq \) and pronounced “is subordinate to” or “is a subDRS of”, is more interesting. By definition, two DRSs stand in this relation, \( \varphi \sqsubseteq \psi \), iff \( \varphi = \psi \) or there is a \( \sqsubset \)-chain of ever growing DRSs connecting \( \varphi \) to \( \psi \). This captures the

\(^{9}\)Though, actually, the DRT counterparts of \( [\varphi \Rightarrow \psi] \leftrightarrow [\neg[\varphi \land \neg\psi]] \) and \( [\varphi \lor \psi] \leftrightarrow [\neg[\neg\varphi \land \neg\psi]] \) are still valid.

\(^{10}\)We are really talking about DRSs as tokens, or specific occurrences of DRSs if you will, because, for example, the DRS \( \langle \{\}, \{\text{know}(x,y)\} \rangle \) as such would be immediately subordinate to two non-overlapping DRSs within (7).
intuitive notion of a DRS box nested inside another. What we are going to need later on is the notion of accessibility that is directly based on this. A DRS \( \varphi \) has access to a DRS \( \psi \) (or “\( \psi \) is accessible to \( \varphi \)”), denoted \( \varphi \prec \psi \), iff \( \varphi \sqsubseteq \psi \) or there is a (superordinate) DRS \( \chi \) containing \( \psi \Rightarrow \varphi \) as a condition. So, a given embedded DRS has access to all DRSs it is nested in, including itself, and in addition, conditionals’ consequents can see their antecedents. Derivatively, a discourse referent or condition is said to be accessible to another referent or condition iff the corresponding DRSs themselves stand in the accessibility relation.

Another important notion is that of free versus bound discourse referents. This is really the exact analogue of predicate logic’s free and bound variables, except that here the only variable binders are the DRS universes: a \( U(\varphi) = \{x, y\} \) as in the main DRS in (7) binds all \( x \) and \( y \)’s occurring in the atomic conditions of \( C(\varphi) \) and in all subordinated DRS conditions. Formulated in terms of accessibility, a discourse referent in an atomic condition is a free variable iff that condition has no access to a DRS universe with that discourse referent. A DRS with no free variables is called closed, and a DRS that, throughout its subDRSs never introduces discourse referents in universes that have access to other DRSs that already use them (in universe or conditions), are called pure. We will henceforth work with pure DRSs only.

3.1.3.2 Semantics

As we’ve said before, DRT comes with a static definition of truth in a model. For the extensional fragment considered so far, the ingredients match those of predicate logic. A model is a pair \( \langle D, I \rangle \) of a domain of individuals and an interpretation function mapping the basic predicates of the language onto appropriate extensions:

\[
\mathfrak{M} = \langle D, I \rangle \text{ is model for a given DRS language fragment}
\]

\(\begin{align*}
a. & \quad D \text{ is a domain of individuals} \\
b. & \quad I \text{ is the interpretation function} \\
& \text{ for every } n > 0, I \text{ maps all } n\text{-place predicates to subsets of } D^n \text{ (and 0-place predicates to truth values, } \in \{0, 1\})
\end{align*}\)

\[\text{\footnote{One sometimes sees definitions of the subDRS notion on which the consequent of a conditional is subordinate to its antecedent. One argument in favor of this is the fact that, as we remarked in footnote 9, } [\varphi \Rightarrow \psi] \leftrightarrow [\neg[\varphi \land \neg\psi]], \text{ the right-hand side of which shows } \psi \text{ obviously embedded into the antecedent } \varphi. \text{ In the current account, we introduce the notion of accessibility to capture this somewhat more inclusive notion of subDRS.}}\]
3.1 Discourse Representation Theory

We will consider very minimalistic DRS language fragments. For instance, since we represent proper names as predicates, we will not consider individual constants, nor function symbols. Arguably, we have no need for \( \iota \)-operators either, since definite descriptions will later be analyzed as presupposition inducers.

As in predicate logic, we need a kind of auxiliary interpretation function for discourse referents/variables. In DRT these are commonly called *embedding functions*, and they are used as interpretation parameter for both DRSs and DRS conditions:

\[
(10) \quad f : \subseteq DRef \rightarrow D \text{ is a proper embedding for a DRS condition } \gamma \text{ if } FV(\gamma) \subseteq \text{Dom}(f)
\]

A proper embedding can either *verify* or *falsify* a DRS condition. This and the notion of DRS verification are defined by simultaneous recursion as follows. First, for a DRS \( \varphi \):

\[
(11) \quad \mathcal{M} \models^f \varphi \text{ iff there is a } g \supseteq f \text{ with } \text{Dom}(g) = \text{Dom}(f) \cup U(\varphi) \text{ (henceforth: } g \supseteq_{U(\varphi)} f) \text{ and for all } \gamma \in C(\varphi): \mathcal{M} \models^g \gamma
\]

Since this depends on the verification of conditions we proceed as follows, starting with atomic conditions:

\[
(12) \quad \begin{align*}
&\text{a. } \mathcal{M} \models^f \xi = \eta \text{ iff } f(\xi) = f(\eta) \\
&\text{b. } \mathcal{M} \models^f \Pi(\xi_1, \ldots, \xi_n) \text{ iff } (f(\xi_1), \ldots, f(\xi_n)) \in \Pi(\Pi)
\end{align*}
\]

Defining truth for complex conditions means we have to rely on DRS verification in (11) again:

\[
(13) \quad \begin{align*}
&\text{a. } \mathcal{M} \models^f \neg \varphi \text{ iff there is no } g \supseteq_{U(\varphi)} f \text{ with } \mathcal{M} \models^g \varphi \\
&\text{b. } \mathcal{M} \models^f \varphi \Rightarrow \psi \text{ iff for all } g \supseteq_{U(\varphi)} f \text{ with } \mathcal{M} \models^g \varphi \text{ there exists a further extension } h \supseteq_{U(\psi)} g \text{ with } \mathcal{M} \models^h \psi \\
&\text{c. } \mathcal{M} \models^f \varphi \lor \psi \text{ iff there is a } g \supseteq_{U(\varphi)} f \text{ with } \mathcal{M} \models^g \varphi \text{ or there is a } h \supseteq_{U(\psi)} f \text{ with } \mathcal{M} \models^h \psi
\end{align*}
\]

For closed DRSs we can now compute whether or not they are true in a model with respect to the empty embedding. If so, we call them simply true in the model: \( \mathcal{M} \models \varphi \). As usual, we often suppress mention of the model and use the following notation, pronounced “the extension(al semantic value) of \( \varphi \)”: \( \llbracket \varphi \rrbracket^f = 1 \text{ iff } \mathcal{M} \models \varphi \) (for conditions, as well as DRSs). This notation will be especially useful when we have intensions as well as extensions, see 3.1.5.
3.1.4 Quantifiers and donkeys

In this section I introduce a small extension to the DRS language needed to capture some aspects of quantification. In 3.1.4.1 we return to our donkey sentences but introduce a new puzzle known as the proportion problem, for which duplex conditions with selective binding will prove a solution. We’ll see how this gives us a nice, general account of quantification in DRT. In 3.1.4.2 I propose to analyze only along the same lines.

3.1.4.1 Duplex conditions

The semantics of DRT is set up in such a way that the classical universal quantification can be defined in terms of a DRS implication where the antecedent’s universe corresponds to the quantified variables, the antecedent’s conditions constitute the restrictor (in generalized quantifier terminology, cf. 2.4.2.2), and the consequent then corresponds to the so-called (nuclear) scope:

\[(14)\] Everybody in the department likes Ligia

\[\begin{array}{c}
\text{x} \\
\text{ligia(x)} \\
\text{y} \\
\text{in_dept(y)} \\
\text{⇒} \\
\text{like(y,x)} \\
\end{array}\]

To see that this is indeed an adequate representation, we must show that the static truth conditions of the DRS in (14) are as represented in the corresponding predicate logical form for (14):

\[\begin{array}{c}
\text{⇒} \\
\exists x [\text{ligia(x)} \land \forall y [\text{in_dept(y)} \rightarrow \text{like(y,x)}]] \\
\end{array}\]

The crucial step in that proof is the application of the semantic rule for \(\Rightarrow\) in the computation of truth of the DRS in (13), call it \(\varphi: \mathfrak{M} \models \varphi\) iff there is a \(f\) with domain \{x\} and \(f(x)\) is Ligia, and for all extensions \(g\) of \(f\) with domain \{x, y\} that map \(y\) to someone in the department, it holds that \(g(y)\) likes \(f(x)\). This is exactly what we get when we compute the the truth conditions of the predicate logical formula with its existential quantifier followed by a universal.

After our sketchy attempt at capturing intersentential anaphora by incorporating a kind of incremental context change, which we will expand upon in 3.2, now we are ready to deal with the second problem of predicate logic: donkeys. The problem, if you recall, was that, on its own, the clause a farmer
owns a donkey is existential, but as soon as we embed it in the antecedent of a conditional the existential quantifiers magically turn into universal ones. In DRT, there’s no problem: according to DRT semantics a discourse referent in a DRS universe is interpreted existentially, unless it occurs as the antecedent of a conditional, in which case it’s interpreted universally:

\[(15) \quad \text{If a farmer owns a donkey, he beats it} \]

This DRS has the correct, universal truth conditions, while at the same time retaining as recognizable subDRSs the representations of each clause on its own, in accordance with the principle of (clausal) compositionality.

This is not the place to launch a comprehensive discussion of the DRT semantics of quantification, but we will need a little more than the classical \(\exists\) and \(\forall\) analogs we have now. One characteristic of our current universal/implication is that it universally quantifies over all discourse referents in the antecedent universe, and it does so unselectively. This means that every farmer has to beat every donkey he owns in order for (15) to be true. Though this may or may not be a reasonable requirement for this particular sentence, it is quite inadequate for the following sentence, which is assigned the same DRS, structurally:

\[(16) \quad \text{Everybody who found a dime in their pocket, put it in the parking meter} \]

Intuitively, it suffices that everybody just uses one of the dimes they have in their pocket. This is an instance of an issue known as the proportion problem, on which much ink has been spilled. I’ll just use it as a reason for introducing selective quantification in DRT.

What we seem to need is a way to specify which discourse referent is universally quantified, and which will remain existential. The following type of condition, called \textit{duplex conditions}, is introduced to facilitate the representation of selective (and unselective) binding:
\[ (17) \]

\[
\begin{array}{c|c}
\hline
x & y \\
\hline
\text{person}(x) & \langle \forall x \rangle \\
\text{dime}(y) & \text{put\_in\_pm}(x, y) \\
\text{find\_in\_pocket}(x, y) & \\
\hline
\end{array}
\]

The syntax and semantics are given in (19), which uses the useful auxiliary notion of a DRS \textit{merge}, \( \oplus \), defined as the pairwise union of referents and conditions:

\[ (18) \]

\[ \varphi \oplus \psi = \langle U(\varphi) \cup U(\psi), C(\varphi) \cup C(\psi) \rangle \]

(19) a. Syntax: if \( \varphi \) and \( \psi \) are DRSs and \( \Xi \subseteq U(\varphi) \), then \( \varphi \langle \forall \Xi \rangle \psi \) is a DRS condition

b. Free variables: \( FV(\varphi \langle \forall \Xi \rangle \psi) = FV(\varphi) \cup (FV(\psi) \setminus U(\varphi)) \)

c. Access: if \( \varphi \langle \forall \Xi \rangle \psi \in C(\chi) \) then \( \varphi \sqsubset \chi \) and \( \psi \sqsubset \chi \), and \( \psi \prec \varphi \prec \chi \)

d. Truth definition: \( \mathcal{M} \models \varphi \langle \forall \Xi \rangle \psi \) iff for all extensions \( g \supseteq \Xi \) verifying \( \varphi \) there is a further extension \( h \supseteq U(\varphi) \setminus g \) that verifies \( \varphi \oplus \psi \)

So, we now have selective binding with universal quantifiers. Another nice thing about duplex representations is that they easily extend to other quantifiers, in fact, the mechanism closely resembles that of generalized quantifiers as discussed before. The quantifier \textit{most} for instance would work like this:

\[ (20) \]

Most farmers who own a donkey beat it

Calculating truth conditions in the style of (19) gives:

\[
[(20)] = 1 \text{ iff most } f \text{ mapping } x \text{ to a farmer with a donkey can be extended to map } y \text{ to a donkey owned and beaten by } x
\]

In effect the truth of (19) thus depend on the ratio between farmers owning at least one donkey and farmers owning and beating at least one donkey. In generalized quantifier theory with lambdas for set abstraction we would
3.1 Discourse Representation Theory

represent these truth conditions as:

(21) \[ \text{MOST} (\lambda x \exists y [\text{farmer}(x) \land \text{donkey}(y) \land \text{own}(x, y)]) \]

This is indeed a plausible reading of the sentence. At least it’s much better than what we’d end up with if we were to have only unselective binding. In that case, we would presumably be counting farmer-donkey pairs, so as to compare the cardinalities of \{\langle a, b \rangle \mid a \text{ is a farmer and } b \text{ is a donkey owned by } a\} and the subset \{\langle a, b \rangle \mid a \text{ is a farmer and } b \text{ is a donkey owned and beaten by } a\}. On such an semantics, one farmer beating all of his 74 donkeys would, counterintuitively, make (20) true even if his 8 colleagues, who each own only one or two donkeys for fun, are very nice to theirs.

3.1.4.2 Only as a quantifier in DRT

Recall our discussion about only in 2.4.2.2. There I speculated that the somewhat simplistic view of only as a generalized quantifier would serve us just fine. But what we can do with generalized quantifiers, we can also do with duplex conditions, so I propose the following DRT representation of only constructions:

(22) Only sadists beat donkeys

Given that proper names and other referential definites are usually represented at the main DRS, we get representations like the following:

(23) Only Colby likes Ligia

The semantics of this duplex condition is based on the observation that only is essentially the dual of all, in the sense that only \( P \)'s are \( Q \) is equivalent
to all Q’s are P. Building this argument reversal into the definition of only requires some extra attention however, for it might well give rise to some some complications with respect to variable binding or its linguistic counterpart, anaphoric binding. But let’s start with an easy example, like (22). It’s clear what we want the duplex condition to mean: all donkey-beaters are sadists, i.e. the duplex condition in (22) should be equivalent to (24):

\[
(24) \quad \begin{array}{c}
\forall x \\
\text{beat_donkeys}(x)
\end{array} \quad \begin{array}{c}
\forall x \\
\text{sadist}(x)
\end{array}
\]

A general definition for deriving this equivalence can take two forms: a syntactic transformation on DRSs or a purely semantic definition. The syntactic rule will have to specify how to turn around the arguments of an arbitrary only condition to get the corresponding universal duplex condition. The only tricky thing there is the treatment of the arguments’ universes in cases where some discourse referents are not quantified over. Consider for instance an only duplex condition with an existentially quantified discourse referent on the right, and see what happens if we represent its intuitive truth conditions with a universal:\[\text{(25) Only a weirdo owns a pet donkey}\]

\[
\begin{array}{c}
\sim \\
\begin{array}{c}
x \\
\text{weirdo}(x)
\end{array} \\
\text{ONLY} \\
\begin{array}{c}
x \\
\text{pet_donkey}(y)
\end{array}
\end{array} \\
\begin{array}{c}
\sim \\
\begin{array}{c}
x y \\
\text{pet_donkey}(y)
\end{array} \\
\text{own}(x,y)
\end{array} \\
\begin{array}{c}
\forall \quad x \\
\text{weirdo}(x)
\end{array}
\]

Conclusion: conditions swap, duplex bound discourse referents remain in their original universe on the left, and the referents on the right move with their conditions to the left.

The only question that remains is what happens to any existential discourse referents in the restrictor of an only condition. And that turns out to be somewhat more difficult question. If they would move with their conditions to the right it would be impossible to bind anaphors in the nuclear

\[\text{12For ease of presentation I have overlooked the fact that we have now created a universally quantified condition where y is bound in both the restrictor and scope of the quantifier whereas it was originally constrained solely within the scope. This will not cause any trouble if we require that the original DRS with only is pure (i.e. introducing fresh variables instead of re-using accessible ones wherever possible).}\]
scope of only with them.

(26) ?Only a man with a donkey beats it

Further research is needed to determine whether or not (and to what extent) this prediction is borne out. The clear impact of focusing on such sentences, leads me to suspect that we need a better understanding of the mechanisms of focusing/backgrounding before attempting an analysis:

(27) a. Only a MAN with a donkey beats it
    b. Only a man with a DONKEY beats it

I have nothing to say about what how exactly intonation alters the logical form of only constructions, or any quantified or unquantified sentence for that matter. And even if I did, it would almost certainly lead us too far astray from our goal. Therefore, at this point I am forced to admit that more has to be said about complex restrictors of the quantifier only and further research is required to investigate whether the position that only can be represented as a quantifier is really tenable in light of these and further data.

At this point, some might be tempted to get rid of the whole idea of only as a quantifier and replace it with a focus based analysis. The reason I will not go down that road has to do with my analysis of bound readings of indexicals (only I did my homework cf. 2.4.3.4, 3.3.3.2) and, ultimately, de re/de se belief reports under only (cf. 2.4.2.2, 3.4.3.2). Further research would be required to determine if these accounts could be adapted to a more focus-based analysis of only. For now, I’ll just stick with the quantificational account and assume some separate information structure module to account for the extra effects witnessed in (27). And as for complex restrictors, the only’s that we will encounter in our study of de re/de se belief reporting will all have non-complex restrictors consisting of no more than a proper name, so let’s not worry about those just yet.

For future reference, we could capture the syntactically described (partial) analysis of only in a simple semantic interpretation clause:

(28) $\mathcal{M} \models^f \varphi (\text{ONLY } \Xi) \psi$ iff for all extensions $g \supseteq \Xi f$ verifying $\psi$ there is a further extension $h \supseteq (U(\varphi) \setminus \Xi) g$ that verifies $\psi \oplus \varphi$

Note that this rule captures a convenient, but somewhat arbitrary, decision to copy unquantified discourse referents in the only restrictor $(U(\varphi) \setminus \Xi)$ to the universal’s restrictor and scope, thus making (26) interpretable, without relying on a particular focus structure.
3.1.5 Intensionality

We are going to be focusing on belief, so we’d better add an intensional dimension to our semantics. In this section we first simply follow the lead of intensional predicate logic. Then we evaluate the belief semantics that gives us, and we investigate how we might want to extend it.

3.1.5.1 Modal DRT syntax and semantics

As in predicate logic, we add modal operators (\(\Box\), \(\text{bel}\), . . . ) to the language and a domain of possible worlds to the model. Then we need to specify the specific interpretations of our modalities, e.g. by giving accessibility relations, and finally we need to add appropriate interpretation clauses to our truth definition. In short:

\[\begin{align*}
\text{Syntax:} & \quad \text{if } \varphi \text{ is a DRS, } \xi \text{ a discourse referent, then } \Box \varphi \text{ and } \text{bel}_\xi \varphi \text{ are DRS conditions} \\
\text{Free variables:} & \quad FV(\Box \varphi) = FV(\varphi) \text{ and } FV(\text{bel}_\xi \varphi) = FV(\varphi) \cup \{\xi\} \\
\text{Access:} & \quad \text{if } \Box \varphi \text{ or } \text{bel}_\xi \varphi \in C(\psi), \text{ then } \varphi \sqsubseteq \psi \\
\text{Model:} & \quad \mathcal{M} = \langle D, W, I, \text{Bel} \rangle^{13}
\end{align*}\]

Up to here, it’s been an entirely straightforward copying exercise. Now we come to the interpretation of the modally extended DRS language in this extended model. Again, there’ll be no surprises. We just have to add an extra interpretation parameter, a possible world, to every clause in the truth definition. There are of course some decisions to be made, but we stay as close as possible to the simple system of 1.A that we’re familiar with. In particular, we choose a purely extensional treatment of discourse referents and equality, so all the intensionality is placed in the interpretation of predicates, i.e. by adding a possible world parameter to I. Let’s spell out some key clauses, keeping in mind that all the embeddings \(f\) here are supposed to be proper embeddings for the terms they are to interpret:

\[\begin{align*}
\text{Truth definition:} & \quad \mathcal{M} \vdash^f \varphi \iff \text{there is a } g \supseteq U(\varphi) \text{ f and for all } \gamma \in C(\varphi): \mathcal{M} \vdash^g \gamma \\
\text{b.} & \quad \mathcal{M} \vdash^f \xi = \eta \iff f(\xi) = f(\eta) \\
\text{c.} & \quad \mathcal{M} \vdash^f \Pi(\xi_1, \ldots, \xi_n) \iff \langle f(\xi_1), \ldots, f(\xi_n) \rangle \in I(\Pi, w) \\
\text{d.} & \quad \mathcal{M} \vdash^f \neg \varphi \iff \text{there is no } g \supseteq U(\varphi) \text{ f with } \mathcal{M} \vdash^g \varphi
\end{align*}\]

\[\text{For simplicity we assume a trivial, universal accessibility for the } \Box \text{ modality, and we reformulate the belief accessibility relations for } \text{bel} \text{ in terms of a belief sets, as we’ve been doing since 1.1.2.1.}\]
And then we need to add two clauses for the new, modal conditions:

\[(31)\]

\[\begin{align*}
  & a. \quad M \models f_w \Box \varphi \text{ iff for all } v \in W: \ M \models f_v \Box \varphi \\
  & b. \quad M \models f_w \text{BEL}_\xi \varphi \text{ iff for all } v \in \text{Bel}(f(\xi), w): \ M \models f_v \varphi
\end{align*}\]

We can now define the very useful notions of extensional and intensional semantic values of an expression, \(\llbracket \ldots \rrbracket_w\) and \(\llbracket \ldots \rrbracket\), respectively:

\[(32)\]

\[\begin{align*}
  & a. \quad \llbracket \varphi \rrbracket_w = 1 \text{ iff } M \models f_w \varphi, \text{ and } 0 \text{ otherwise.} \\
  & b. \quad \llbracket \varphi \rrbracket = \left\{ w \in W \mid \llbracket \varphi \rrbracket_w = 1 \right\} \\
  & c. \quad \llbracket \varphi \rrbracket = \llbracket \varphi \rrbracket^0
\end{align*}\]

For the sake of uniformity we sometimes denote the extensions and intensions of discourse referents and predicates (given by \(f\) and \(I\) directly) with \(\llbracket\)-brackets as well. For DRSs and conditions we usually pronounce \(\llbracket \varphi \rrbracket\) as “the (classical) proposition expressed by \(\varphi\)”, for predicates we stick to “intension”, and for discourse referents we speak of “reference”.

### 3.1.5.2 Beyond intensionality: belief

So now we have the familiar notion of a proposition as a set of possible worlds, and a modal semantics of \(\text{BEL}\) to go with it. We can reformulate the interpretation clause of \(\text{BEL}\) somewhat to bring out the underlying idea that belief is a relation between an individual and a set of worlds, i.e. a proposition, in line with the predicate logical definitions of 1.1.2 and 1.A:

\[(33)\]

\[M \models f_w \text{BEL}_\xi \varphi \text{ iff } \llbracket \varphi \rrbracket^f \supseteq \text{Bel}(f(\xi), w)\]  

[\(\equiv(31b)\)]

Note that this serves only as a first approximation, because, as we’ve seen in the previous chapters, there’s more to belief and report semantics. We’ve seen how omniscience problems and essential indexicality pose severe threats to the simple, intensional belief semantics sketched here. One way to improve on these points was to add a dimension of context-dependence to the semantics in order to properly handle indexicality, and then, in the enhanced, two-dimensional semantics, introduce more fine-grained semantic

---

\(^{14}\)In some textbooks, extensional semantic values of DRSs are sets of strictly verifying embeddings, i.e. the minimal embeddings that verify the condition set: \(\llbracket \varphi \rrbracket_w^e = \{ g \mid g \supseteq_U f \text{ and for all } \psi \in C(\varphi): M \models g \psi \}\). Intensions, then, are sets of pairs \(\langle g, w \rangle\), \(g\) strictly verifying \(\varphi\) in \(w\). This setup facilitates an elegant reformulation of the basic truth definition, but it’s equivalent to the one given here. We like to think of \(\llbracket \varphi \rrbracket\) as denoting the proposition expressed by \(\varphi\), and thus we’re led to the definitions in (32).
Chapter 3. A semantics of attitude reports

objects (characters, diagonal properties) for belief. We will not go down that road—at least, not yet—for two reasons, which we shall dwell on for the remainder of this section. Both have to do with natural alternatives to two-dimensionalism offered by the dynamic and representational nature of the DRT framework itself. It remains to be seen if these are viable alternatives for analyzing indexicality and belief, so let’s take a look.

First, we have switched to a dynamic semantics primarily in order to deal with context dependence, so we’ll examine a truly dynamic treatment of indexical context dependence before falling back on Kaplan’s static contexts and characters. This native DRT account treats all definites, including definite descriptions, anaphoric pronouns, indexicals and proper names, as presupposition inducing expressions. In 3.2 we take up presupposition theory and its uniform treatment of definites. We’ll see in 3.3 however that this purely dynamic account of indexicals and proper names turns out to be unsatisfactory unless we add some Kaplanian elements to it, which will eventually result in a hybrid analysis superior to either framework. Indeed this framework employs semantic context parameters alongside dynamic discourse contexts. Still later, in 3.3.4, we use this to refine our modal semantics of the BEL operator somewhat. Note however that the real work lies in the representation of natural language reports, more specifically, we’re ultimately interested in the way English belief reporting sentences can convey particular de re and de se meanings, depending on context and syntactic structure. We return to this difficult issue in 3.4. What I’m trying to get across here is that all this talk about a good semantics of BEL is not directly about the semantics of English belief reporting constructions. It has more to do with the underlying belief semantics, which, as I’ve been arguing in the previous chapters, should be treated as independently as possible.

The second prima facie reason not to go two-dimensional already, is that DRT’s representationalism offers a native account of structured belief. In 1.1.2.3 we saw how simple possible worlds analyses of attitudes were plagued with logical omniscience, and how adding structure to the belief objects could help reduce the number of unwanted inferences. The ultimate structured belief object would be a sentence, or at least some formal, syntactic object, structured like a natural language sentence, supposedly modeling a language of thought. On such an analysis, believing something means having the corresponding sentence of thought stored in your belief box. We criticized this analysis for a variety of different reasons already, but it does effectively render logical omniscience and essential indexicality harmless. Now, DRT may be seen as providing a superior structured belief object: the DRS. This would be totally compatible with the original mentalist inspiration behind DRT that saw DRSs as representing what happens in the mind of an interpreter.
It is only natural then to define someone believing something, as having the mental state represented by the corresponding DRS. In such a model, we’d replace the belief set with a DRS capturing the individual’s mental state. It’s obvious that this is essentially still a sententialist approach to omniscience, and as such it’s pretty successful. For example, the DRSs for *Ligia is a philosopher and 37 is prime and Ligia is not not a philosopher* [= (13), p. 13] are very different and we may well assume that a structured mental state (whether a DRS or a set of DRSs) can contain the first but not the second. Furthermore, things may be set up in such a way that we can do justice to the more difficult belief puzzles involving co-referential *res* terms in belief (historical examples include Frege’s *Hesperus/Phosphorus*, Quine’s *Ortcutt* and Kripke’s *Pierre in London/Londres*) also including essential indexicals (*Lingens, Heimson, . . .*). The formal machinery involved is quite intricate, so I refer the interested reader to Zeevat (1984; 1986), Asher (1986; 1987; 1993) and Kamp (1985; 1990). Furthermore, the resulting framework of DRS belief objects, with *internal* and *external* anchors, has even been successfully applied to the notorious belief puzzles of intentional identity (Geach’s (1967) *Hob/Nob*).

To sum up, (i) the cognitive science roots of DRT provide independent motivation for DRSs as representations of mental states, and (ii) the resulting structured belief account has been worked out and successfully applied to even the most obstinate belief puzzles that the philosophy of language has to offer. In addition, (iii) the account beats basic sententialism in that its highly structured belief objects can also be interpreted in a model: DRSs are only an intermediate stage of interpretation, somewhere in between language and the world, i.e. DRS representations are still connected to the external world by a real Tarskian truth definition. In other words, DRT can take into account both a *narrow* and a *wide* perspective on meaning and mental states (cf. 1.2.4.4 and 3.1.2.2).

Although in principle I’m all for the use of DRSs as structured belief objects, I will not develop the idea here any further. This is essentially for the same, mostly practical, reasons I did not follow up on any of the omniscience counter-measures listed in 1.1.2.3: I am primarily interested in report semantics and thus will choose the simplest, most transparent semantics of belief possible. Unfortunately, despite the fact that it is in some sense the obvious way to go in DRT, taking DRSs themselves as belief objects in the semantics makes for a much more complicated theory than does a simple modal, or even two-dimensional, alternative. Besides, I want to be able to compare my results in *de re* report semantics with the static state of the art alternatives of chapter 2, all of which use some kind of doxastic possible worlds framework. I will here abandon the DRS-belief account in favor of a
possible worlds account. Whether this will require one or two dimensions of intensionality remains to be seen after we’ve first developed an account of indexicals as definites, i.e. as presuppositions, in the next section.

3.2 Presupposition

Presupposition theories have become an important and integral part of the various forms of dynamic semantics. Here we discuss what presuppositions are (3.2.1), and how we can incorporate them into DRT (3.2.2). The proposed presuppositional extension of DRT is what inspired the two-stage interpretation model that we’ve been talking about and that we’ll continue to work with. In 3.2.3 we apply our theory to belief reports and get a nice analysis of the \textit{de dicto/de re} distinction.

3.2.1 Triggers, satisfaction, accommodation, and projection

Utterances often presuppose a certain amount of information, i.e. information that has to be assumed as given for the utterance to make sense. Take (34) for instance:

(34) I talked to the bearded logician

Uttering this sentence out of the blue will most likely raise some eyebrows, unless the discourse context somehow already provides a specific logician. Therefore, we say that (34) presupposes the existence of a bearded logician. The first thing to note is that presuppositions are triggered by specific linguistic constructions, called presupposition triggers. In this case it’s the definite description that triggers the existential presupposition. By contrast, indefinites usually lack presuppositional force, i.e. \textit{I talked to a bearded logician} does not presuppose, but simply asserts the existence of a logician with a beard: if there were no such logicians, (34) would be nonsensical whereas the indefinite variant would just be false. Other examples of presupposition triggers include factives, (35a), it-clefts, (35b), particles like \textit{too, also and again}, (35c), and so-called aspectual verbs (\textit{start, stop}), (35d).

(35) a. Ligia knows that Colby likes her
   \hspace{1cm} \Rightarrow \hspace{1cm} \text{Colby likes Ligia}

b. It was a \textsc{logician} I talked to
   \hspace{1cm} \Rightarrow \hspace{1cm} \text{I talked to someone}

c. I also talked to \textsc{colby}
3.2 Presupposition

I use $\Rightarrow$ to denote the presupposition normally associated with an utterance of the sentence, i.e. to denote the information taken for granted when one utters such a sentence. It is important to note that the term presupposition is also used to denote the material that is conventionally invoked by presupposition triggers.\(^\text{15}\) We shall see cases where a sentence constituent triggers a presupposition in this last sense, even though an utterance of the whole sentence does not presuppose anything in the first sense.

Characteristic of presuppositions is also their tendency to survive any kind of embedding. The classic examples of this involve embeddings under negation, but it holds more generally for any kind of embedding, including e.g. connectives and intensional operators:

\(\text{(36) }\)

a. Ligia doesn’t know that Colby likes her
   $\Rightarrow$ Colby likes Ligia
b. If I also talked to Colby, I must have been quite drunk
   $\Rightarrow$ I talked to someone else
c. Some people may not believe that he quit smoking yesterday
   $\Rightarrow$ He used to smoke

The combination of (35a) and (36a) suffices to show that presupposition is not a special kind of (classical) inference, for something entailed by a statement and by its negation would have to be a tautology.\(^\text{16}\) But then why is it that normally utterances of the sentences above do seem to imply their presuppositions? The answer lies in the context dependence of utterance and presupposition. Presuppositions, we said, impose certain requirements on a discourse, to the effect that a presupposing utterance demands a discourse context in which its presuppositions are fulfilled. Apparently, (35a) and (36a) are felicitous in the same contexts, viz. contexts where it is understood that Colby likes Ligia. Thus it is no surprise that any felicitous discourse containing these sentences entails that Colby likes Ligia.

As it turns out, things are not as clear-cut as that. There are many cases of an arguably presupposing sentence uttered felicitously in a context where the presupposition is not already common ground. For instance, we can easily imagine a discourse where Ligia uses (35d) to explain Colby’s annoying

\(^{15}\)For this second sense I later introduce the marker $\partial$.

\(^{16}\)This argument only holds for classical two-valued logic, which is precisely why for some time, three or more valued logics were de rigueur in presupposition theory.
behavior to a friend of hers that never met Colby before and therefore had no idea he used to be a smoker. In such a case, the presupposition triggered by the aspectual verb *quit* in Ligia’s utterance is not contextually given, but new information, for the hearer at least. This use of presuppositions, to convey new information, is called *accommodation*, following Lewis (1979b). One popular way to think of accommodation is as a repair strategy: (35d) triggers a presupposition that Colby used to smoke, so the context sketched in the scenario above is defective; accommodation of the presupposition amounts to an enrichment of the context with the presupposition’s content; after this fix, the presupposition is trivially verified, the sentence is felicitous and the discourse entails that Colby used to smoke. The status of accommodation within the theory of presupposition is still hotly debated; for some, in the tradition of Heim (1983b), such a repair strategy has no place in a compositional, semantic theory of presupposition and assertion, but should be relegated to a truly pragmatic level of interpretation concerned with what people actually do with words in particular situations (cf. Beaver 2001). In the school of thought underlying the current work on the other hand, accommodation is analyzed on a par with contextual (or local) givenness of presuppositions. This second type of analysis thus allows a unified formal analysis of accommodation, which on the Heimian view ended up in the pragmatic wastebasket. The cost of this is that the formalization will be non-compositional in a very real sense, and moreover, will rely essentially on the representational level offered by DRT’s originally mentalistic framework.

But there’s one more basic fact to know about presupposition before we go into formalization. The third characteristic trait of presuppositions is that they are *defeasible* in the sense that a sentence that would be presuppositional on its own, may lose its presuppositional character under certain embeddings. A case in point is (37), which does not require a context where the speaker is known to have talked to anyone, and what is more, the utterance doesn’t even entail that she talked to anyone:

(37) If indeed I spoke to anyone at all last night, it was a logician I talked to.

The consequent of this conditional, i.e. (35b), would normally presuppose that I talked to someone, but if we embed it in a conditional that relativizes it to the information otherwise presupposed, the resulting utterance no longer presupposes anything. The phenomenon that a given clause’s presupposition often survives embedding, but is sometimes defeated is called *presupposition projection*. This used to be known as the projection problem, but with dynamic semantics soon came a satisfactory analysis. The main idea was
that dynamically interpreting a conditional like (37), we first, temporarily, increment the discourse context with the antecedent’s content. This gives a so-called local context, which serves as the background against which the consequent is evaluated. In a case like (37), the antecedent adds enough information to satisfy the presupposition triggered in the consequent. The presupposition is satisfied locally, which explains the observation that the sentence as a whole no longer poses restrictions on the context. The fact that conditionals don’t entail their antecedents then explains why this sentence, even if embedded in a felicitous discourse, doesn’t entail that I talked to someone.

Let me finish with some terminological remarks. Firstly, in order not to confuse presupposition with implicature, a different type of pragmatic informational content often characterized by its ‘cancelability’, I will refrain from talking about presupposition cancellation in favor of presuppositions being ‘defeated’. Secondly, when talking about presupposition projection on a pre-theoretical level, I will speak of justification or verification of a presupposition, instead of satisfaction, because the latter is too easily associated with a particular theory of presuppositions, viz. Heim’s (1983b) Satisfaction Theory. In fact, the theory I favor, van der Sandt’s (1992), cashes out the general idea of presuppositions being defeated in terms of a stronger notion called presupposition binding.

3.2.2 Presupposition as anaphora in DRT

In this subsection I will detail what I consider the most natural analysis of presupposition within the DRT framework: van der Sandt’s (1992) theory of Presupposition as Anaphora (PA), also known as the Binding Theory of Presupposition (Geurts 1999). As I briefly pointed out above, its main contemporary rival is Heim’s (1983b) Satisfaction Theory of presupposition, which is based on Karttunen (1974) and Stalnaker (1974). For lack of space I will forgo an in depth comparison of the two theories, in fact I will only develop PA, mostly because it’s the obvious choice once we’ve already decided to go with DRT; the Satisfaction Theory is designed specifically to meet the demands of compositionality and non-representationalism imposed by FCS and DPL. The interested reader is referred to Beaver (2001) and Geurts (1999) for (not unbiased) overviews and comparisons of the modern theories of presupposition.
3.2.2.1 Two-stage interpretation

So let’s focus on the DRT+PA analysis of the phenomena described in 3.2.1. As discussed before, nowadays we like to think of DRT as a two-stage interpretation process that separates the construction of a preliminary DRSs from the incrementation of the context.\(^\text{17}\) I have little to say here about the first stage, except that we will require that preliminary DRS be compositionally derivable from the sentence’s surface structure. In practice, I cannot always provide the exact derivations, let alone a fully general system to parse a certain fragment of English and map it to the corresponding preliminary structures. In other words, the exact formulation of this so-called *construction algorithm* is not my concern here, but I trust that in my example DRSs it will always be clear that such a compositional derivation of the preliminary structure could be given straightforwardly.

On to stage two then: the *resolution algorithm*. This algorithm takes a preliminary and a context DRS as input and gives a new context DRS as output. The resulting two-stage interpretation of a multi-sentence discourse can illustrated schematically as follows (construction algorithm denoted \textbf{Prel}, resolution abbreviated \textbf{Res}):

\begin{equation}
\text{(38)} \quad \text{Let } \langle \sigma_1, \ldots, \sigma_n \rangle \text{ be a discourse, then } \varphi_i, \text{ the representation of the initial segment } \langle \sigma_1, \ldots, \sigma_i \rangle, \text{ is given by:}
\begin{align*}
\text{a. } & \quad \varphi_0 \text{ represents the context at the start of the discourse} \\
\text{b. } & \quad \varphi_{i+1} = \textbf{Res}(\varphi_i \oplus \textbf{Prel}(\sigma_{i+1}))
\end{align*}
\end{equation}

But what exactly is the place of presupposition in all this? As we saw, presuppositions are triggered by specific phrases or constructions in a sentence. A preliminary DRS of a sentence will need to mark that it has triggered presuppositions in order for the resolution algorithm to check whether these are indeed satisfied in the context, or whether they should be accommodated. Given the possibility of presuppositional information being already provided somewhere earlier in the same sentence, as in (37), it is important that preliminary DRSs encode precisely where (in the \textless hierarchy of the DRS) the

\(^{17}\text{Originally this distinction was not made, as witness the standard DRT textbook of Kamp & Reyle (1993) in which dynamic interpretation proceeds by adding to the context DRS a syntactic parse tree of a sentence and then gradually turning that tree into DRS conditions, resolving context dependencies such as presuppositions and anaphora on the fly. Given that DRT was supposed to model incremental information growth in the mind of a hearer, this may be the most natural way to model dynamic interpretation. However, an empirically adequate and elegant treatment of presuppositions turned out to work out better if we separate construction and resolution. Therefore, the new textbook by Kamp \textit{et al.} (1999) does incorporate a two-stage interpretation strategy, even at the cost of sacrificing some of the psychological appeal of the earlier, fully incremental implementation.}
presupposition is triggered. We add a special operator \( \partial \) (symbolizing partiality) to the language to mark presuppositions in preliminary DRSs. The content of a presupposition is represented as a DRS, prefixed with this \( \partial \).

In box notation we like to use the more visual dashed boxes instead of the \( \partial \) prefix. In the compositional construction of a preliminary DRS such a presupposition box is inserted at the point where the construction algorithm parsed its trigger.

Let’s look at an example:

\[
\text{(39) } \text{The logician is asleep}
\]

\[
\begin{array}{c}
\text{sleep}(x) \\
\text{logician}(x)
\end{array}
\]

In this preliminary DRS we see that the sentence asserts no more than that an individual \( x \) is asleep. Who this individual is, is to be determined by resolution of the presupposition triggered by the definite NP. The dynamic interpretation of this representation may be paraphrased as: find a representation of a logician in the context, say \( z \); set \( x = z \); and add the information that \( x \) sleeps. Dashed boxes thus represent a request to search for a suitable antecedent to identify their presupposed discourse referent with. The metaphor of ‘looking for an antecedent’ brings to mind the name of the framework, Presupposition as Anaphora. Indeed, the basic idea is that the same representation and resolution methods can be applied to anaphora and presupposition alike. We’ll look at resolution examples below, but a preliminary representation of a sentence with a anaphoric pronoun would look like this:

\[
\text{(40) } \text{She is asleep}
\]

\[
\begin{array}{c}
\text{sleep}(x) \\
\text{female}(x)
\end{array}
\]

Since proper names and indexical pronouns are context dependent definites as well, we might want to represent them as presuppositions/anaphors too. They certainly share the main trait of presuppositionality, viz. that they
require a contextually given antecedent for the discourse to be acceptable: if I am aware that nobody in my audience knows who Colby is, I shouldn’t want to enter the conversation with (41a), just as I cannot use the definite subjects of (39) or (40) if I know their antecedents are not common ground. With pure indexicals, the story is essentially the same, but note that we need their antecedents to be given in the actual, external context. This raises a delicate issue which we will discuss at length in 3.3. For now, let’s just treat names and indexicals as regular presupposition triggers, keeping in mind that these representations will be refined in 3.3:

(41)  

a. Colby is asleep

\[
\begin{array}{c}
\text{sleep}(x) \\
\text{x} \\
\text{colby}(x)
\end{array}
\]

b. I am asleep

\[
\begin{array}{c}
\text{sleep}(x) \\
\text{x} \\
\text{speaker}(x)
\end{array}
\]

Now, let’s look at embedded triggers and their presuppositions. In view of local binding possibilities it is important to place the presupposition boxes at the right level of embedding, i.e. the level at which they are triggered (as already dictated by compositionality). Unfortunately, the examples of (35) all involve presupposed entities such as events and propositions which we are not equipped to deal with yet. So let’s stick with definites:

(42)  

Most people Ligia knows believe that she is dating the bearded logician

\[
\begin{array}{c}
\text{sleep}(x) \\
\text{x} \\
\text{speaker}(x)
\end{array}
\]
3.2 Presupposition

The first trigger, a proper name, is syntactically embedded in the restrictor of a quantifier, the other two are even further embedded, inside a belief complement inside the quantifier’s scope. By compositionality, the presuppositions they give rise to must be located accordingly within the hierarchical structure of the preliminary DRS, as they are in the DRS in (42).

3.2.2.2 Binding presuppositions

Now that we know, at least roughly, how to construct preliminary representations, let’s bring the context into the picture and consider some example resolutions. Let’s assume a context where Ligia is already known. Of course, the existence of Ligia is not the only thing that is common ground, but let’s simplify and represent as little as possible:

\[(43)\]

\[
\begin{array}{c}
  x \\
  ligia(x)
\end{array}
\]

Now I utter (44) (preliminary DRS given below):

\[(44)\] Ligia is friends with a handsome linguist and a bearded logician

\[
\begin{array}{c}
  \begin{array}{c}
    y \\
    z
  \end{array} \\
  friends(w, y) \\
  handsome(y) linguist(y) \\
  friends(w, z) \\
  bearded(z) logician(z) \\
  \text{Most}
\end{array}
\]
The indefinites are represented as simply introducing new discourse referents and conditions to the preliminary DRS. The proper name, of course, triggers a presupposition. Merging (44) with the context, (43), gives:

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ligia(x)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>friends(w,y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>handsome(y) linguist(y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>friends(w,z)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bearded(z) logician(z)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this big DRS we now need to find an antecedent for the presupposition. The first constraint on resolution is a semantic one:

MATCH The content of a presupposition must *semantically match* the interpretation of the antecedent

Let me illustrate matching on the current example. Applied to (45), it basically means I can take x as the antecedent for w, since what x refers to is clearly something that fits the content of the presupposition. More precisely, for all g with x ∈ Dom(g) that verify the non-presuppositional part of (45), the presuppositional DRS is verified by the extension of g that maps w to g(x). I will not yet offer a fully general and exact formulation of what it means for a presupposition to match with an antecedent, but I aim here to capture the idea that binding can only occur when the antecedent’s content *entails* the presupposition’s content.\footnote{Determining the ‘antecedent’s content’ in a multiply embedded DRS containing variables bound higher up, is a serious technical challenge. Let’s see how van der Sandt (1992) deals with it. Note first that he weakens MATCH to a consistency check: we bind a presupposition to an antecedent if the content associated with that antecedent does not conflict with the presupposition’s content. I consider the resulting notion of binding a mixture of binding and accommodation, as to be discussed at the end of 3.2.2.3, but the point here is that van der Sandt depends on a way of collecting all information currently associated with a candidate antecedent for checking what he calls *local consistency* and *local informativity*. Unfortunately he does not properly define these notions, as witness Beaver’s (1997) criticism: Van der Sandt’s formulations of informativity and consistency constraints seem to involve a notion of local entailment of sub-DRSs, although I am not aware of such a notion ever having been formalized. \[(Beaver 1997:980)\]}

\[18\] I choose this rather strong entailment...
3.2 Presupposition

model for now because I think it comes closest to the pre-theoretical ideas of dynamic presupposition justification and filtering, and also to Heimian satisfaction. Later on, we’ll discuss a weakening of the notion of matching, merely requiring consistency of the presupposition with the content associated with the antecedent. I postpone this point of divergence between PA and the Satisfaction Theory until after we have introduced accommodation in PA.

Once we have identified a semantically matching antecedent, we bind the presupposition to it. Formally this consists in a unification of the discourse referents involved, in this case that comes down to a substitution $[w \rightarrow x]$, and deletion of the presupposition. The idea behind this deletion is that, once we have identified $x$ and $w$, the presupposition’s content is justified by the context: it constitutes information already entailed by the context above it and is thus superfluous. The result is a closed and pure DRS, representing the old context updated with the information conveyed by the first sentence:

$\begin{array}{|c|c|c|}
\hline
x & y & z \\
\hline
ligia(x) & friends(x, y) & handsome(y) linguisit(y) \\
friends(x, z) & bearded(z) logician(z) \\
\hline
\end{array}$

This DRS is again truth-conditionally interpretable: it is true if the model provides a person called Ligia and two friends of hers, one a handsome linguist, and one a bearded logician.

In this new context we would expect that (42) can be uttered felicitously. Let’s see how we can derive that. First, we need to construct the preliminary DRS, which we did in (42) already, and then we merge that with our context:

Fortunately, in later work on PA, more sensible formulations of the so-called local constraints have come up. Cf. Kamp’s (2001) still sketchy notion of an “extended” local context, which consists of the information in the local context representation itself together with all information located at those contexts to which the local context is subordinate” (p.8). For a real definition, I refer the reader to Bos’s (2003:203-4) implementation of the relation supersub that is determined by walking through the DRS along the accessibility paths while gathering the relevant pieces of information. Interestingly, Geurts (1999) seems to think we can make do with only ‘global’ versions of the constraints in question, more concretely, he reduces them to constraints grouped under our PRAGMATICS (the output must be informative and consistent), which we will encounter below. Beaver (1997) also proposes a constraint on finished output DRSs, but one that checks informativity and consistency for all subDRSs.
Here we have embedded presuppositions looking for antecedents and then a second constraint on binding kicks in:

ACCESS an antecedent must be accessible from the (sub)DRS where the presupposition is triggered.

Alternatively, we could try and derive this constraint from an apparently more general constraint:

CLOSURE the output DRS must be closed

Under certain standard assumptions on Prel, closure implies access, since the introduction of a presupposition always creates a preliminary DRS with free variables, viz. the discourse referents that occur in the presupposition universes (cf. (39)-(42)). To bind those free discourse referents we unify them with already bound ones, as exemplified in (45)-(46). But this will only close the open DRS conditions if these conditions have access to

\[\text{Prel} \text{ is well-formed,}\]

\[\text{a DRS } \varphi \text{ contains a presupposition } \partial \psi \in C(\varphi) \text{ then it also contains an atomic condition } \Pi(\xi_1, \ldots, \xi_n) \in C(\varphi) \text{ containing at least one } \xi_i \text{ from the presuppositional universe } U(\psi).\]

Once we allow Prel to form preliminary structures violating this well-formedness restriction, closure can be achieved by binding presuppositions to inaccessible antecedents. van der Sandt & Huitink (2003) exploit this feature in their analysis of again.
the antecedent universe, otherwise, if we were to bind the presupposition in a position inaccessible (e.g. strictly subordinated) to the trigger, the free variables left behind at the trigger location would remain free, even after unification with the bound antecedent. Because I’m not sure I really want to enforce the aforementioned ‘standard assumptions on Prel’ I’ll just take closure and access as two separate constraints.

Combining the two main constraints on binding gives: presuppositions look for a semantically matching antecedent that is accessible from the trigger location. Let’s apply this to our example. The first presupposition was the one triggered by the proper name in the quantifier’s restrictor. As defined in 3.1.4, a restrictor does not have access to its quantifier’s scope, just as a conditional’s antecedent has no access to its consequent, only the other way around. The only accessible DRSs thus are the restrictor itself and of course the main DRS. In the restrictor we find $u$, but there’s nothing that makes this $u$ satisfy the presupposed content of being Ligia, so we turn to the global context, where indeed we find a Ligia to bind to, so we substitute $x/v$.

Then we take on the pronoun’s presupposition. All four subDRS are accessible from this position, but the first two have empty universes, and the third only has $u$, who need not be female at all. Though our simplified initial context representation makes no explicit mention of it, we may well assume that it’s commonly known that Ligia is a woman, if only because it’s a girl’s name, so again $x$ seems a plausible candidate to bind to. The bearded logician, obviously, is out, because of the beard, but why not take the handsome linguist? Well, it depends, in certain situations, e.g. if you’re convinced that ‘Ligia’ must refer to a man, we might well bind she to the linguist $y$; as a general rule, just pick whatever makes the most sense in the given situation. Let’s turn this into a third constraint on binding:

PRAGMATICS pick the resolution candidate that will eventually yield a pragmatically optimal output interpretation

The term ‘pragmatically optimal’ here hides a competition of a number of (mostly Gricean) constraints such as coherence, non-redundancy and relevance, which I will only refer to on a very informal level.\(^{20}\) We should be careful about the order in which to apply these constraints, the semantic/formal ones should outrank the pragmatics (but for the system so far we

\(^{20}\) A straightforward formalization of the competing constraints idea may be in the framework of Optimality Theory, which has been successfully applied to both semantics (de Hoop & Hendriks 2001) and pragmatics (Blutner 2000). We could use OT not only within the family of constraints dubbed PRAGMATICS, but also to the whole constraint based definition of the binding algorithm sketched above. The most natural order of the constraints introduced so far would be: \{CLOSURE, ACCESS\} > MATCH > PRAGMATICS.
get part of that ordering for free, since an open DRS cannot even compete in the race for pragmatic optimality). A proper order may also muffle the combinatorial explosion that we get from computing possible resolution combinations to get interpretable output candidates whose felicity we can then compare according to PRAGMATICs.

In any case, in the current example, we bind $s$ to $x$, and then we go and look for an antecedent for the last presupposition, *the bearded logician*, which can only be $z$. The final output of $\text{Res}$ is the following, closed and proper DRS:

$$
\begin{array}{c|c|c|c|c|c|c}
\hline
x & y & z \\
\hline
\text{ligia}(x) & \text{friends}(x,y) & \text{handsome}(y) & \text{linguist}(y) & \text{friends}(x,z) & \text{bearded}(z) & \text{logicain}(z) \\
\hline
\end{array}
$$

Before we look at accommodation, we take a look at a local verification example where a presupposition triggered in a conditional’s consequent is filtered out because it gets bound by asserted material in the conditional’s antecedent:

$$
(49) \quad \text{If I did talk to anyone, it must have been a logician I talked to.}
$$

The sentence DRS for this one will involves two presuppositions for each of the two first person pronouns, and one for the it-cleft, to the effect that I talked to someone. So, the second $I$’s presupposition is actually embedded in the it-cleft’s. In cases of multiply embedded presuppositions, $\text{Res}$ will take the most deeply embedded presupposition first, if only because it is difficult to check MATCH if the presupposition to be bound carries a stack of further unresolved presuppositions with it. So after merging with a minimal context, with nothing more than a speaker, we resolve the $I$’s and we have:
As for the it-cleft’s presupposition, the conditional’s antecedent is accessible and introduces a perfectly matching discourse individual, i.e. someone the speaker talked to, so we need not look any further and bind there:

We thus get a perfectly good (closed, pure, pragmatically sound) DRS, without demanding anything of the context but a speaker, exactly as described in 3.2.1.

Considering many more examples of local, or rather intermediate, binding, a tendency to bind close to the trigger has been observed in cases where there are various accessible, matching candidate antecedents. We thus formulate a new resolution constraint, ranked on a par with the pragmatic ones:

**LOCAL** bind a presupposition as close to the trigger as possible

For computational implementation this may be especially useful as it implies that a presupposition starts its search for an antecedent in the DRS where it’s triggered; if it finds no suitable antecedent there, it gradually works its way down the so-called *projection path* (the complete chain of accessible DRSs) until, if all else fails, it arrives at the global DRS, where it is either bound or reports failure.

---

21This is a case where we need a better formalization of ‘the content associated with a discourse referent’ in the definition of the MATCH constraint, as discussed earlier in footnote 18.
3.2.2.3 Accommodation

With all these constraints on binding, it may well happen that no suitable antecedent for a presupposition is found. In such a case we can resort to our aforementioned repair strategy of accommodation. To capture the fact that accommodation really is a repair strategy and thus only applies after all binding options have been ruled out, we could introduce another constraint on Res:

\textsc{repair} avoid accommodation

As presented in 3.2.1, accommodation means we fix the context in such a way that it will trivially satisfy/bind the presupposition, viz. by adding the presupposed information to it. In DRT this comes down to merging the presupposition with an appropriate DRS accessible from the trigger location. We thus reformulate the \textsc{access} constraint on binding, to include accommodation:

\textsc{access} accommodation and binding sites must be accessible from the trigger location

Depending on where on the projection path this merge occurs, we speak of global, intermediate or local accommodation.

Let’s consider examples of each of these types of accommodation. And at the same time, let me introduce another bit of notation, $\triangleright$, for denoting resolution processes, as opposed to $\triangleright$ which signifies the (underspecified) representation of a sentence in isolation, i.e. a PrelDRS (or lf, in other frameworks).

(52) I think Ligia likes the logician

\[
\begin{array}{|c|c|}
\hline
x & y \\
\hline
\text{speaker}(x) & \text{ligia}(y) \\
\hline
\end{array}
\]

$\triangleright$

\[
\begin{array}{|c|}
\hline
\text{BEL}_{x} \\
\hline
\text{like}(y,z) \\
\hline
\text{z} \\
\hline
\text{logician}(z) \\
\hline
\end{array}
\]

\[
\leadsto
\begin{array}{|c|c|}
\hline
x & y \\
\hline
\text{speaker}(x) & \text{ligia}(y) \\
\hline
\end{array}
\]

\[
\leadsto
\begin{array}{|c|}
\hline
\text{BEL}_{x} \\
\hline
\text{like}(y,z) \\
\hline
\text{z} \\
\hline
\text{logician}(z) \\
\hline
\end{array}
\]
3.2 Presupposition

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>speaker(x)</td>
<td>ligia(y)</td>
<td>logician(z)</td>
</tr>
<tr>
<td>BEL_x</td>
<td>like(y,z)</td>
<td></td>
</tr>
</tbody>
</table>

I’ve represented the input context, the context merged with the preliminary DRS, and, after some trivial presupposition bindings, the output. In the last step, binding according to our constraints was not an option, so we resorted to global accommodation. With the constraints formulated so far, a weaker, local accommodation would have been possible as well. Several accommodation constraints have been formulated to derive a preference for global accommodation for this example:

GLOBAL  accommodate as far away from the trigger as possible

This is the dual of LOCAL, which said that presuppositions want to bind as close to the trigger as possible. Another constraint that works for this example, but gives different predictions in others is:

STRENGTH  choose the accommodation site that maximizes the output’s logical strength

This formulation uses the notion of \( \varphi \) being stronger than \( \psi \) iff \( \varphi \) asymmetrically entails \( \psi \). Thus, it can be viewed as a specific instance of the general pragmatic maxim “be informative!” Without further argument I will opt to have GLOBAL replace, or at least outrank, STRENGTH. For an empirically informed defense of GLOBAL over STRENGTH I refer the reader to Geurts (2000).

However, in some cases, the global DRS, though trivially accessible, is not available for accommodation (or binding) due to the structure of the DRSs and presuppositions involved. We want our outputs to be properly interpretable (i.e. closed), as specified by our earlier constraint CLOSURE. When CLOSURE prevents the accommodation process from reaching the global DRS we speak of trapping. In (53) the presupposition triggered by his car is trapped in the restrictor by the quantified Germans: 22

(53) Every German loves his car

22I skipped the part where the possessive pronoun triggers a presupposition that gets bound, intermediately, by the x in the restrictor. Idem for (55).
The resulting reading *every German who has a car loves it*, is quite plausible, but others have defended a stronger reading, *every German owns a car and loves it*, which we would get by local accommodation of the presupposition:

Since we chose GLOBAL over STRENGTH, local accommodation is dispreferred. There is a heated debate between satisfaction theorists and PA people over such examples, for it is one of the issues on which the two really make divergent empirical predictions.

Finally, contextual, pragmatic factors may influence accommodation preferences, so that even purely local accommodation becomes possible:

(55)  
[Why is Colby so sad?]
[You mean that guy? Well, I don’t know...] Perhaps his girlfriend broke up with him, or maybe his dog died.
Here, binding is out, but there’s no trapping, and the global and intermediate positions do give closed, interpretable results. The local accommodation reading however is the one that makes the most sense, given standard conversational maxims applied to the context sketched. The global accommodation output for instance would imply that (it is common ground that) Colby had a girlfriend and a dog, which is contradicted by the fact that the second speaker doesn’t even remember for sure who Colby is. In other words, pragmatics applies to accommodation as well as binding and is ranked at least as high as global.

A final constraint is needed to prevent accommodation of pronouns, because it is almost impossible that a sentence like (56a) can mean there is a man who likes to talk, by accommodation of the presupposition triggered by the pronoun. It has been hypothesized that this has to do with the lack of specific content in a pronominal presupposition. Adding more content to the presupposition makes the sentence much easier, as witness the acceptability of the examples below, all uttered in a context where the subject term cannot be bound to a contextually familiar discourse referent:

(56)  

a. *He likes to talk

b. ?My friend likes to talk
Chapter 3. A semantics of attitude reports

c. My best friend Colby from high school likes to talk

Presumably, the rich content of the bigger definite in (56) makes it more eligible for accommodation because it allows the hearer to construct a reasonably accurate picture of the kind of individual the speaker has in mind. As a constraint:

**SPECIFICITY** do not accommodate presuppositions low on informational content

This applies to indexical (1st and 2nd) pronouns as well, but for such pronouns, accommodation does not even come into play usually, because in any sane conversational context it’s quite clear who addresses whom. Note that there seems to be a connection with rigidity/direct reference, though proper names seem to be able to accommodate on occasion (Geurts 1997). We return to the issue of rigidity in DRT in section 3.3.

We have now discussed the major pragmasemantic factors that play a role in computing optimal resolutions. But until now, we have looked at binding and accommodation as separate strategies for resolving presuppositions, and because of that we have overlooked one important resolution option: the binding/accommodation mix. So far, binding meant that the presupposed discourse referent and conditions were fully ‘absorbed’ into the antecedent, which could only happen if there was a *match* between them, i.e. if the antecedent’s content entailed the presupposition. Accommodation on the other hand meant that the whole presupposition was added as new information to a DRS. What if the presupposed content only *partially matches* the antecedent’s, i.e. it does not really follow from it, but it is not contradictory either (Krahmer & van Deemter 1998). A simple example:

(57) A famous professor visited the department. She gave a talk.

<table>
<thead>
<tr>
<th>x y</th>
</tr>
</thead>
<tbody>
<tr>
<td>famous(x)</td>
</tr>
<tr>
<td>prof(x)</td>
</tr>
<tr>
<td>dept(y)</td>
</tr>
<tr>
<td>visit(x,y)</td>
</tr>
</tbody>
</table>

What to do with the pronoun’s presupposition? **PRAGMATICS, REPAIR and SPECIFICITY** would select binding to x over accommodation, but, as it stands,
MATCH does not allow that since the context does not specify that the famous professor is indeed female.\textsuperscript{23} By far the most intuitive reading of the discourse is one where \( z \) binds to \( x \) but accommodates its content:

\[
\begin{array}{c|c}
 & x & y \\
 \hline
\text{famous}(x) & & \\
\text{prof}(x) & & \\
\text{dept}(y) & & \\
\text{visit}(x,y) & & \\
\text{give\_talk}(x) & & \\
\text{female}(x) & & \\
\end{array}
\]

The presupposition is essentially bound, but part of its content is accommodated into the context, i.e. we learn from the second sentence that this famous professor mentioned in the first was a woman. For the purpose of formalization we classify such resolutions as binding, since the presupposed discourse referent gets identified with an existing one, but it does necessitate two small adaptations (built into most versions of PA from the start (van der Sandt 1992; Geurts 1999)). First, binding itself now consists in the unification of discourse referents plus the addition of the presupposition’s condition set to the binding site. In cases of full matching, this results in semantically superfluous conditions, which, by way of notational convention, will simply remain invisible in our box notations. Second, MATCH must be weakened to allow partial matching. Perhaps, one way to go about this would be to weaken it all the way down to a mere consistency check and trust that the other constraints, especially PRAGMATICS, will provide enough further restrictions on binding possibilities. A further refinement to be considered is that the accommodation of conditions is subject to the general accommodation constraints REPAIR and SPECIFICITY. Unfortunately I do not have time to go into these matters in detail, but in the following we will often make use of partially matched binding/accommodation.

To sum up, we have tried to cast van der Sandt’s (1992) presupposition resolution mechanism in the form of a competition of (partial) output candidates according to a set of constraints on binding and accommodation. In other words, we’ve made some first steps towards an implementation of PA’s Res in OT. It’s just a rough sketch though, and certainly not ready for computational implementation (see Bos (2003) for that). In fact we haven’t even been very specific about the exact ranking of the constraints.\textsuperscript{24} I hope I’ve

\textsuperscript{23}For this example, the other constraints are irrelevant.

\textsuperscript{24}I think we’d get a good approximation of van der Sandt’s proposal by using a standard OT, taking preliminary DRSs as input and generating full DRSs as output candidates,
nonetheless managed to clarify the ideas behind DRT+PA, because in the following we rely heavily on the two-stage formulation of DRT, with a Prel and a Res algorithm essentially operating as discussed here. There will be some minor adjustments necessitated by the introduction of layers in 3.3 and second order variables, but then we’re all ready for our analysis of de re/de se reporting. But first, in the next section, let’s look at what we get if we simply merge our modal analysis of belief and PA into a simple account of attitude reports, and why that fails.

### 3.2.3 Presuppositions in attitude reports

We’ve seen in the previous section how presuppositions tend to escape any kind of embedding, including embedding under a bel operator. We’ve left it more or less open whether we can really keep the simple modal semantics of bel conditions, or whether something more sophisticated is required, but let’s just represent “x believes that...” as belx and worry about the underlying belief semantics later. In 3.2.3.1 I’ll show how treating definites as presuppositions leads to a new and improved version of the old scopal analysis of the de dicto/de re distinction (2.2.1). Though it does get some things right, it also suffers some of the same defects as its Russellian predecessor, i.e. in de re readings we get truly singular propositions and therefore we run into problems with Ortcutt scenarios (1.2.2). In 3.2.3.2 we thus follow the reasoning of 2.2.2 and introduce res descriptions into belief. Finally, we’ll want to extend our account to incorporate the insights of 2.3.1 and allow treatment of the somewhat subtler de re/de se distinction. To do that however we need to move from propositions to properties, or, from worlds to contexts. But for a proper introduction of semantic contexts to the DRT framework we will have to wait until 3.3 adds a layer for indexicals with accompanying interpretation. We’ll pick up belief reports again in 3.3.4.

In this section we also return to the examples discussed throughout chapter 2, because we want to test our theorizing against the concrete sentences and contexts that we have been using to test other theories of belief reporting. The test cases in question are 2.1’s sentences (1)-(3) and the contexts described in scenes A, B, and C. To apply our dynamic account, bear in mind that we shouldn’t simply represent the complete stories as DRSs and take those as input contexts. That would mean that all that information using the soft constraints described above and ranking them as follows:

\[
\begin{align*}
(i) & \quad \left\{ \text{closure, access, match} \right\} < \left\{ \text{pragmatics, local, repair} \right\} < \text{specificity} < \text{global} < \text{strength}
\end{align*}
\]
3.2 Presupposition

is already common ground between the reporter and the audience, in which case the reports in question would become redundant. Of course, we may assume that the reporter has full knowledge of what’s going on with Ellsworth, but not all of that is common ground yet. For the most part we take the scene description to establish a possible world, in which to evaluate the predicted output DRS for truth or falsity. So, in order to adapt our paradigm to the dynamic setting, we must split the stories into a background context or common ground, modeled as a DRS, and a factual context of evaluation, modeled as a possible world.

Concretely: for the interpretation of all our reports we need the background contexts of A, B, and C to provide someone called Ellsworth. For A and C this should suffice to make most of the utterances interpretable, but, following our earlier discussion of the crucial judgments of (2) in connection with ‘mistaken self-identity’, we had better assume that it is commonly known that Ellsworth is watching someone called Arata Suggs on TV and that that Arata is actually herself. For future reference, let me suggest the following dynamic and evaluation contexts for A, B, and C:25

\[
\begin{array}{|c|}
\hline
x \\
\hline
\text{ellsworth}(x) \\
\hline
\end{array}
\]

(58) \( \varphi_A: \)

\( c_A: \) Ellsworth is the prettiest, though she doesn’t believe that. In fact, she is a bit disappointed because she believes the prettiest will win, saying: “The pretty ones always win, I don’t stand a chance.” Meanwhile, Ellsworth is known to some as Arata Suggs.  

\[
\begin{array}{|c|}
\hline
x \\
\hline
\text{ellsworth}(x) \\
\text{arata}(x) \\
\text{see_on_tv}(x, x) \\
\hline
\end{array}
\]

(59) \( \varphi_B: \)

\( c_B: \) Ellsworth is still the prettiest, but she doesn’t think that matters anymore. Arata is a nickname name for Ellsworth, though she thinks this Arata, the person she’s watching on TV, drunk, is someone else, someone better, so she says: “She’s really good, this Arata on TV, I think she’ll win, not me”  

\[
\begin{array}{|c|}
\hline
x \\
\hline
\text{ellsworth}(x) \\
\hline
\end{array}
\]

(60) \( \varphi_C: \)

25The DRS representations of (58)-(60) can be taken as highlighting only the minimally required and highly salient knowledge present in the discourse context.
Chapter 3. A semantics of attitude reports

**cC:** Ellsworth, full of self-confidence, thinks: “Looks don’t matter, I’ll definitely win” For compatibility: Arata=the prettiest=Ellsworth.

[cf. C, p. 74]

For completeness, here are the reports we want to evaluate, and the table of our intuitive judgments, whose columns indicate the truth values of the sentences uttered in the background of A, B, or C, and evaluated in the corresponding world.

(61) a. Ellsworth Kimmel believes pretty girls always win \[= (1a), \text{p. 72} \]
b. Ellsworth believes the prettiest girl will win \[= (2), \text{p. 72} \]
c. Ellsworth believes of Arata Suggs that she’ll win \[= (1b), \text{p. 72} \]
d. Ellsworth believes she’s going to win \[= (3), \text{p. 73} \]
e. Ellsworth believes to be on the winning side \[= (1c), \text{p. 72} \]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>(61a)</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(61b)</td>
<td>1</td>
<td>?1</td>
<td>?1</td>
</tr>
<tr>
<td>(61c)</td>
<td>0</td>
<td>1</td>
<td>?1</td>
</tr>
<tr>
<td>(61d)</td>
<td>0</td>
<td>?1</td>
<td>1</td>
</tr>
<tr>
<td>(61e)</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.1: Intuitions

3.2.3.1 Presupposition as scope in *de dicto/de re* reports

DRT+PA currently predicts that presuppositions triggered inside a belief report generally project out of that belief, in order to be bound or accommodated at a higher level. Let’s see what this means exactly by applying it to the paradigm *de dicto/de re* reports (61a)-(61d).

First (61a), uttered in \(\varphi_A\). Since there’s no presupposition triggered inside the belief, we just get a straightforward translation\(^{26}\) into a DRS by \textbf{Prel} alone—apart from the trivially resolved proper name in the main clause, which we will henceforth no longer represent explicitly.

(62) Ellsworth Kimmel believes pretty girls always win \[= (61a) \]

\(^{26}\)I will not go into the question how bare plural + *always* gives a universal reading. Note also that a universal quantification over events would be closer to the truth about (62), but I don’t want to go into event semantics either.
The truth conditions of the output correspond to the *de dicto* reading that we want: in all of Ellsworth’s belief worlds pretty girls always win. In fact, translating the DRS into predicate logic we’ll see the exact same If as assigned in chapter 2, to wit (4), p. 76. Putting the sentence in the context of $\varphi_B$ doesn’t change anything, so, as expected, the outputs are true only in $c_A$, because in the other worlds Ellsworth does not believe looks matter.

Let’s move on to an allegedly ambiguous report, (61b). First, we add the preliminary structure to the minimal context, $\varphi_A$. We see a presupposition triggered inside a belief box:

(63) Ellsworth believes the prettiest girl will win $[=61b]$
Chapter 3. A semantics of attitude reports

The context is so minimal indeed that we have to accommodate the prettiest’s presupposition. Global is preferred over local (by \textit{GLOBAL}): 

\begin{align*}
\begin{array}{|c|}
\hline
x \\
\hline
\text{ellsworth}(x) \\
\hline
\end{array} & \leadsto \\
\begin{array}{|c|}
\hline
\text{BEL}_{x} \\
\hline
\text{win}(y) \\
\text{y} \\
\text{prettiest}(y) \\
\hline
\end{array}
\end{align*}

This output truth-conditionally corresponds to the wide scope if presented in 2.2.1, i.e. (6b), p. 76. In the computation of a truthful embedding we first evaluate who’s prettiest and then ask whether that individual wins in Ellsworth’s belief worlds. It doesn’t matter who Ellsworth considers pretty, it’s about who actually \textit{is} the prettiest (or rather, whom we (the reporter and her audience) consider the prettiest). The content of \textit{the prettiest} is thus used only to fix a certain individual that the belief is about, which makes this a \textit{de re} logical form. Since we assumed that Ellsworth is the prettiest in all of \(c_A\) to \(c_C\), the belief is \textit{de re} about Ellsworth, which makes it true in \(c_B\) and \(c_C\), because there Ellsworth verbalizes her belief in the form of a sentence about herself (though she is unaware of that aboutness in \(c_B\)).

We can also imagine that due to all manner of contextual factors not represented explicitly here, this reading doesn’t quite fit in to the wider conversational context. Perhaps because it’s obviously false, or because it’s somehow redundant. In such a case we could invoke a violation of \textit{pragmatics} to enforce local accommodation:
Thus, we get a reading equivalent to (6a), p. 76. This narrow scope/local accommodation output obviously corresponds to a de dicto reading: in all of Ellsworth’s belief worlds there’s a prettiest girl and that one wins. Note that it doesn’t matter who’s actually pretty, we only evaluate who’s prettiest according to Ellsworth. As expected this DRS is true only in the first scenario, $c_A$.

The same resolutions are possible in $\varphi_B$ and $\varphi_C$. In none of these does the context provide matching, accessible antecedents for the presupposition to bind to. If it did, globally or locally, we’d of course immediately bind to them without even considering accommodation. The net result will still be the same outputs as (64) or (65), but now there would be only one candidate output, the one given by binding to the nearest antecedent. For instance, if we were to assume that it was common ground that Ellsworth herself is the

---

27To prove the equivalences (64)$\equiv$(6b) and (65)$\equiv$(6a) you’ll need the fact that the predicate prettiest already verifies uniqueness, so we can indeed leave out the $\iota$ constructions. We’ll see examples later on where uniqueness is not part of the predicate but where we nonetheless choose not to use a $\iota$ operator. The idea behind this simplification in DRT+PA is that the uniqueness that is often inherent in the use of a definite is captured at a meta-level, viz. by the fact that, as presuppositions, they want to project out and bind to an already given, hence definite, discourse individual. It is therefore not the case that we replace the contribution of the predicate logical $\iota$ simply with an existential quantifier.

28How, you ask, can we get local binding with this particular sentence? How can the context make available a pretty girl within the belief introduced by the current sentence? The answer involves something called modal subordination. I cannot go into details, but consider this discourse:

(i) Ellsworth believes there’s always one contestant who’s considered to be the prettiest, and one who’s the smartest. [Furthermore,] Ellsworth believes the prettiest will win.

The first sentence creates a belief box for Ellsworth with a prettiest and a smartest. I propose that the second, our (63), doesn’t introduce a second belief box, but adds to the already given one. Somehow, the parallel construction with repetition of believes licenses this. Incorporating this into our theory requires some adaptation of the ‘merge-and-resolve’ component of our interpretation strategy. Note, by the way, that this kind of modal subordination is nowhere near as mysterious as the Hob/Nob problem of intentional identity, for that involves belief of different individuals somehow linked together.
prettiest, we’d predict only the \textit{de re} reading. I think this prediction is more or less borne out, i.e. I expect for instance the following natural language inference to be valid:

(66) Ellsworth is the prettiest. She believes the prettiest will win. Therefore, she believes she herself will win.

We see how adding information to the background may influence the outcome of the resolution process (by blocking a possible accommodation site, or blocking accommodation completely). This is one big advantage of the flexible dynamics of DRT+PA.

The other, related, advantage of the DRT+PA analysis over the old scope account is that we have gotten rid of a syntactic ambiguity. In DRT+PA we can make do with a single ‘logical form’, a compositionally derived preliminary DRS, viz. the one in (63). The distinct truth conditions are derived \textit{pragmatically}, by putting the preliminary DRS in context and applying $\text{Res}$. In other words, instead of a stipulative If ambiguity, we have underspecification that is resolved in context.

Let’s also take a quick look at the \textit{de re} construction of (61b). If pressed, I’d propose the following preliminary structure and resolutions:

(67) Ellsworth believes of Arata Suggs that she’ll win \hfill [= (61c)]

The hearer should accommodate the existence of someone called Arata Suggs and then the sentence ascribes to Ellsworth a \textit{de re} belief about that individual (so the hearer should also infer that Ellsworth is properly acquainted
with this individual). Since in our evaluation contexts, Arata is Ellsworth herself, the ascribed de re belief is about Ellsworth herself, which is verified only by contexts \( c_B \) and \( c_C \).

In the context of \( \varphi_B \) the nickname Arata for Ellsworth is saliently common knowledge so we can bind instead of accommodate:

\[
\varphi_B = \begin{array}{c|c|c}
\text{x} & \text{ellsworth(x)} & \sim \\
\hline
\text{arata(x)} & \text{see_on_tv(x,x)} & \sim \\
\end{array}
\]

Again, a de re belief of Ellsworth about herself, so true in \( c_B \) and \( c_C \).

With the coreferential pronoun report (61d) we get outputs similar to (68), independent of the input context. The embedded presupposition triggered by she cannot be bound locally, because there is no local antecedent, so it will be bound globally, by Ellsworth. Accommodation is absolutely no option here (by repair and specificity). So we only get one possible output, which is, moreover, truth-conditionally equivalent to (68).

\[
(69) \quad \text{Ellsworth believes that she’s going to win} \quad [=(61d)]
\]
Chapter 3. A semantics of attitude reports

\[ \varphi_A = \varphi_C \approx \varphi_B \leadsto \]

\[
\begin{array}{c|c|c}
\text{prelim.} & \text{context} & \text{resolution}^29 \\
\hline
(61a) & \varphi_A, \varphi_C & \text{glob.acc.} \\
(61b) & \varphi_B & \text{glob.bind.} \\
(61c) & \varphi_A, \varphi_C & \text{glob.acc.} \\
(61d) & \varphi_B & \text{glob.bind.} \\
\hline
\end{array}
\]

Table 3.2: Predictions of DRT+PA’s scopal analysis of \textit{de dicto/de re}

The first thing to note is that our predictions match our semantic intuitions (table 3.1), and also the predictions of the simple scope account (table 2.2). The big plus here is that we don’t rely on syntactic ambiguity. In a sense we moved the \textit{de dicto/de re} distinction into semantics/pragmatics. Unfortunately, the account, like its lf-scopal predecessor, cannot handle Ortcutt cases, as we shall see in 3.2.3.2, nor can it do full justice to the \textit{de se} mode, as it involves too coarse-grained (singular) propositional belief objects, as to be discussed in 3.2.3.3.

3.2.3.2 \textit{De re as de dicto} in DRT

The scope account did well on our Ellsworth related test cases, but failed with respect to a Quinian Ortcutt scenario. The same goes for the dynamification proposed in the preceding section. The reason is that in both cases we eventually model \textit{res} with directly referential expressions (variables bound from the outside). This gives singular propositions for sure, but, as we saw

\[ \text{This of course refers to the resolution of the embedded subject.} \]
in 2.2.2, it’s too strong a model of intentionality. Let’s just go through the crucial example (from 1.1.3.2) again:

(70) Janell is a big fan of Professor Ortcutt’s work. One day, when refereeing for a conference, she rejects a certain anonymous abstract, saying: “This guy’s totally stupid”. Unbeknownst to her, the abstract was in fact Ortcutt’s.
   a. Janell believes Ortcutt is dumb
   b. Janell believes Ortcutt is brilliant

The puzzle is that both sentences are fine, and in fact both report a de re belief of Janell’s. On a wide scope analysis of de re however, we can conclude from (70a) and (70b) together that Janell must have a contradictory belief state.

Now, to apply our dynamically enhanced version, we must first divide up the story into a background (DRS) and an evaluation context (possible world). I propose to put the two distinct ways Janell is acquainted with Ortcutt in the background representation. The reason for this is that I am convinced that, out of the blue, a discourse containing both (70a) and (70b) is indeed incoherent. Though the decision will not have any consequences yet, the idea of utilizing that contextually given information will eventually form the basis of my own analysis in 3.4.

\[
\begin{array}{c|c|c}
 x & y \\
 \hline
 \text{janell}(x) & \text{ortcutt}(y) & \text{famous}(y) \\
 \text{admire}(x,y) & \text{referee}(x,y) \\
\end{array}
\]

\(\varphi(70)\); Janell: “Ortcutt is great, but that guy who wrote abstract #042 is dumb”. Ortcutt wrote #042.

What do we predict then? Say, we were to utter the two reports in succession—a bit of an oversimplification because in reality I’d expect to need some intervening statements (preferably contributing part of the contextual background \(\varphi(70)\)) to make it really acceptable as a discourse.\(^{30}\)

\(\text{\textsuperscript{30}Here’s a more realistic discourse based on (70) containing both reports:}\)

(i) Did you hear who turned down professor Ortcutt for the conference? It was this PhD. student called Janell Ventress, maybe you’ve heard the name. Did you see the comments? Apparently she thinks Ortcutt’s pretty dumb. Funny thing, when I met her the other day she said she’d read all his books and thinks Ortcutt’s
The output here is inconsistent, for it implies that there is this famous Ortcutt being refereed by Janell, represented as $y$, and in all of Janell’s belief alternatives this individual is dumb, but at the same time, he’s brilliant in all of her belief worlds.\footnote{If we take the idea that proper names are just presupposition triggers, with content, seriously, we should expect a local, or even global accommodation option. We could for}
3.2 Presupposition

Exactly as in static semantics, Ortcutt cases show the need for taking descriptions into account. In 2.2.2 we argued for Kaplan’s (1969) way of going about this, essentially turning *de re* belief into *de dicto* belief by replacing the *res* with a vivid name of the *res* for the believer. We can simply plug the Kaplanian theory into our DRT framework, and get the following derivations for (70):

---

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>janell(x)</td>
<td>ortcutt(y) famous(y)</td>
</tr>
<tr>
<td>admire(x,y)</td>
<td>referee(x,y)</td>
</tr>
</tbody>
</table>

(i) $\text{BEL}_x$

- $\text{dumb}(y)$

(ii) $\text{BEL}_x$

- $\text{brilliant}(z)$

- ortcutt(z)

---

Unfortunately this doesn’t work in general, because the name used in the report to pick out the *res* may well be distinct from any description known to the attitude holder. For instance, if my audience consists of professor Ortcutt’s close friends, I could have used her first name, Deandre, in my reports, even if Janell doesn’t know Ortcutt’s first name. When informing Ortcutt directly of Janell’s funny misidentifications of her, I’d use the appropriate indexical (*Janell thinks you are dumb/brilliant*), which under a local accommodation would give demonstrably wrong truth conditions (at least if Janell is not also present at my talk with Ortcutt). Additionally, it’s not even strictly speaking a solution to the puzzle as set out in chapter 1, because one of the argument’s premises was that both beliefs ascribed were *de re*. The local accommodation resolution makes the second one completely *de dicto*. 
This output is easily verified since the context supplies two appropriate individual concepts to map \( d \) and \( d' \) to. As shown in 2.2.2, if we apply this to the Ellsworth examples and scenarios we’d get predictions matching those of the scopal account. But, without even looking very closely at the exact predictions, there are some disadvantages. A minor technical consideration is that we need to enrich the language with intensional discourse referents \((d)\) and predicates that apply to them \((R)\). More seriously, we have to give up the scopal unification of de dicto/de re provided by DRT+PA, because we’d still have the usual in situ, or narrow scope, preliminary structures for de dicto beliefs, but we need more than a simple wide scope resolution for de re, apparently we need the properly quantified and restricted vivid names already given in the de re preliminary structures, thus reintroducing a syntactic ambiguity between de dicto and de re:

(74) Ellsworth believes the prettiest girl will win \[=\text{(61b)}\]
3.2 Presupposition

Obviously a highly unsatisfactory state of affairs about which more should be said, and if it weren’t for lack of space, I would certainly try my hand at a proper unification of *de dicto* and *de re* incorporating the Kaplanian insights. As it is, we’re focusing on *de re/de se* issues and besides, we’re about to replace our vivid names with *de se* sensitive acquaintance relations anyway, so, for what it’s worth, let me just sketch very briefly the kind of unification I envisage, borrowing ideas from Kraut (1983) and Heim (1992).

The starting point would be something like the quantified vivid name structure, (74b). My first step would be to weaken the acquaintance/vividness notion built into the semantics of $R ((7), p. 78)$ so as to allow individuals to be ‘acquainted’ with concepts instead of with just real objects. Next, I’d extend the resolution algorithm to allow certain descriptive presuppositions to bind to (and accommodate as) individual concepts, based on the idea that the use of the definite *the prettiest* may be justified by the givenness of a prettiest individual, but also by the contextual availability of the concept of ‘being the prettiest’. This should do the trick: the presupposition triggered by *the prettiest* in the main DRS can be bound to an actual prettiest known to Ellsworth, or it can give rise to the accommodation of the *the prettiest* concept. In the first case, resolution proceeds as in (73), i.e. *de re*. In the second, we can just equate $d$ to the accommodated concept and get the right *de dicto* reading, disguised as a *de re* reading about a concept, to wit, the concept derived from the subject term surfacing in the report complement.
3.2.3.3 From vivid names to acquaintance relations

Let’s leave *de dicto* reporting and the much discussed *de dicto/de re* interface behind and start focusing on the lower half of our test data in (61): *de re* and *de se*. In 1.2.3 we have argued that essential indexicality necessitates a more fine grained analysis of belief, involving self-ascription of properties and centered contexts. The belief expressed by Ellsworth’s “I am going to win” is not merely a *de re* belief of Ellsworth about Ellsworth, under some vivid name, but rather, a pure *de se* belief. Pure *de se* beliefs cannot be modeled in the still propositional vivid name framework, for that we need Cresswell & von Stechow’s (1982) extension dubbed *relational attitudes*. Now, Ellsworth’s first person belief does fall under a general *de re/de se* definition: it’s a relational belief of Ellsworth about Ellsworth under the acquaintance relation of identity. In 2.3 these insights were straightforwardly incorporated in a report semantics, and applied to Ortcutt, and the remaining *de re/de se* example sentences, repeated here as (61c), (61d) and (61e). We got nice results, especially for (61d), the one with the anaphoric pronoun, which was assigned the right truth conditions without the need for a *de re/de se* ambiguity on which Chierchia (1989) has to rely. The problem lies in the unambiguously pure *de se* PRO construction, (61e), which we would have to assign the same general relational lf.

All of this is directly applicable to DRT after we’ve extended the DRS language to include e.g. $\lambda$-abstraction (to turn DRSs into properties), and higher-order discourse referents:\[32\]

(75) Ellsworth believes that she’s going to win

[$\neg (61d)$]

\[
\begin{array}{c|c|c}
\hline
\text{R} & \text{R(x,y)} \\
\hline
\triangle & \text{ellsworth}(x), \text{fem}(y) \\
\hline
\text{BEL}_x \lambda u \quad \text{R(u,v)} \\
\hline
\text{win(v)} \\
\end{array}
\]

\[32\text{One apparent difference with the static analogue is due to the fact that we don’t want } \iota \text{ operators in DRSs. Since we should really restrict R to range only over suitably vivid acquaintance relations, the definition of which comes with a clause for uniqueness in the second argument (cf. footnote 37, p. 50), this difference will be inconsequential.}\]
3.2 Presupposition

\[ \phi_A = \phi_C \approx \phi_B \rightarrow \]

Our output is true in both \( c_B \) and \( c_C \). For \( c_B \) we can verify the existence of \( \mathbf{R} \) with the contextually available relation of seeing someone on TV. For the pure \( de \ se \) context \( c_C \), we can simply take the egocentric acquaintance relation of equality as an \( \mathbf{R} \).

Great. But what about the unambiguously 1st person \( de \ se \) report, (61e)? If we assume a PRO as embedded subject and further assume that that PRO is, presuppositionally or syntactically, bound by the subject, Ellsworth, we get exactly the same resolutions as for the co-referential pronoun example above, cf. 2.3.1. Since our intuitions clearly showed a contrast with respect to \( c_B \), the mistaken identity scenario, we obviously need more. Incorporating Chierchia’s (1989) main insight we developed a ‘hybrid’ or ‘intermediate’ theory that left the analysis of pronouns reports intact, but directly assigns a non-relational property self-ascription to PRO reports (2.4.1.3-2.4.1.4):

(76) Ellsworth believes PRO to be on the winning side \([= (61e)]\)

Note how this is truth conditionally equivalent to the output in (75) if we replace \( \mathbf{R} \) in the latter with \( = \), which is as it should be (cf. table 1.1, p. 54).

All in all, we have reached a stage where we can do justice to all our test intuitions, as demonstrated in table 3.3, plus the additional Ortcutt example. Are we done then? Well there’s a couple of things that still need our attention. We may distinguish a number of theoretical and empirical shortcomings.

Let’s start with the theoretical qualms: First of all, we still need to look into the semantically 1-dimensional, presuppositional treatment of proper names and indexicals, especially given that we’ve sneaked in Kaplanian con-
texts through the back door, viz. in the semantics of the self-ascriptive, property-sensitive BEL operator (cf. 1.2.3, (43), p. 46). Secondly, the syntactic ambiguity distinguishing de dicto and de re readings of a definite description based report at the preliminary DRS level, and the somewhat similar fact that pure de se reports encoded via PRO have a dedicated de se structure in which there is no presuppositional element corresponding to an embedded subject, whereas the co-referential pronoun report complement does trigger a presupposition for its pronominal subject. Note that both outputs are nonetheless similar in that replacing the quantified acquaintance relation with equality in the second gives us the first. We’re going to exploit this similarity in our unification of de re/de se in 3.4. Next, in 2.3.2 we argued for contextualization of the acquaintance relation, but it remained a rather vague idea whose actual implementation was left to an unspecified pragmatic module. With our move to dynamic semantics we finally have a good grip on context dependence and context change, but, as table 3.3 shows, we’re not taking advantage of it all, in fact, even less so than in our enhanced scopal account (which turned out to be insufficient for different reasons however), cf. table 3.2. To my mind, the crucial argument for the contextualization of acquaintance was that:

one can analyze [(70a) (=Janell believes that Ortcutt is dumb)] as true in one context but false in another. I feel that an analysis which simply characterizes [(70a)] and [(70b)] as both true does not do justice to my intuitions. [(Abusch 1997:9, fn.9)]

And the same applies, perhaps even more so, to the de re/de se example (75): apart from the somehow unmarked pure de se interpretation that we get in C, we’ve agreed that an impure reading is needed, but it seems to me that one should only be available if the discourse context provides a particular and salient non-egocentric way of picking out Ellsworth for Ellsworth, like seeing someone on TV in B. For such cases I totally agree with Abusch that it’s insufficient to say that sentence (75) is simply true in both scenarios B

<table>
<thead>
<tr>
<th></th>
<th>prelim. context resolution</th>
<th>output</th>
<th>cA</th>
<th>cB</th>
<th>cC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(61a)</td>
<td>n.a.</td>
<td>(62)</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(61b)</td>
<td>n.a.</td>
<td>(62)</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(61c)</td>
<td>n.a.</td>
<td>(62)</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(61d)</td>
<td>n.a.</td>
<td>(62)</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(61e)</td>
<td>n.a.</td>
<td>(62)</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.3: Predictions of DRT+PA’s relational hybrid
and C, as the quantified relational account currently predicts. In 3.4 I’ll take advantage of our dynamic framework to implement a proper contextualization of de re/de se reports to do justice to this intuition. The concomitant split of contexts into a dynamic, discourse context (modeled as a DRS), and a static, semantic context (modeled as a centered possible world) will play a crucial role in tackling particular examples. Last but not least in the list of theoretical objections is the questionable compositionality of Prel given the ‘wide scope property’ of the relational (and the vivid name) accounts, as discussed in detail for the static variants in 2.3.3.

Each of these theoretical issues comes with a corresponding empirical manifestation, so let’s list those in order. First, the presuppositional treatment of referential terms makes a couple of interesting predictions, some unorthodox but deserving of further attention, others just plain false. The unorthodox predictions involve the possibilities of non-global binding and accommodation of indexicals; the wrong ones involve the examples used in the Kripke/Kaplan arguments against descriptivity (cf. 1.2.4.2) and are due to a collapse of the important distinction between indexicals and their descriptive counterparts, e.g. I vs. the speaker, or this vs. the object pointed at. Both of these points will be taken up in 3.3 where I couple the presuppositional account of rigidity with a two-dimensional semantics.

Second, the de dicto/de re ambiguity, about which I have nothing more to say, and the hybrid analysis of de se that postulates a dedicated pure de se LF/Prel(iminary )DRS for infinitival reports. Perhaps our analysis doesn’t count as a real unification of de re/de se reports of the form NP believes that NP VP, since infinitivals now have no subject, but if there were no empirical objections, it would be quite plausible to say that infinitivals are special in that their semantic representation does not involve a presuppositional embedded subject or quantified acquaintance relation. This line of reasoning was already debunked in 2.4.3, by the introduction of Schlenker’s (2003) Amharic data (93). Translated into DRT the point there was that the dedicated pure de se PrelDRS, (76), may be plausible for pure de se reports with PRO and perhaps even for reports with logophors, but it is insufficient for Amharic I and the numerous other examples of shifted indexicality, because in such examples there is clearly a referential/presuppositional subject term, with an independently well established semantic content. Amharic I, for instance, is a first person pronoun, so, pending some technicalities to be addressed by the 2-dimensional extension in 3.3, it triggers a presupposition looking for a speaker. Indeed, in many contexts Amharic I behaves exactly as its English counterpart (i.e. as a real first person indexical rather than as an anaphoric third person pronoun, or a special purpose de se element like he*, PRO or a logophor, cf. 2.4.3.1), but when embedded in belief reports
it has a reading that flaunts Kaplan’s Prohibition of Monsters by referring to the ascribee, rather than the person who’s actually uttering the report. In 3.4 I follow Schlenker in analyzing Amharic I as a real indexical, but one that somehow gets evaluated with respect to the belief alternative contexts, rather than the actual world. It is not at all trivial to cast this idea into the general form provided by the acquaintance based framework and some extensions are needed. In any case, the way this is cashed out in my dynamic, pragmasemantic, presuppositional, relational framework is radically different from Schlenker’s system (2.4.3.3), or the rival system of von Stechow (2002) (2.4.3.5). Both of these rival analyses rely heavily on a morphosyntactic theory of agreement (2.4.3.4), which my own account seeks to minimize in favor of a flexible pragmatic presupposition resolution mechanism. The next step is to treat PRO and logophors semantically just as Amharic I, which solves the problem of the dedicated pure de se PrelDRS and clears the way for a full unification of de re/de se.

The third issue involved the missed opportunity of the acquaintance relation’s context dependence, which is scheduled for repair in 3.4. However, this immediately raises questions about the data involving quantified belief reports, which, in 2.4.2, were shown to cause problems for a straightforward contextualization of the relational framework. In 3.4.3 we show how a minor extension to the acquaintance resolution framework can handle all these problematic examples.

The fourth and final theoretical worry, viz. the one about Prel being non-compositional, has no direct empirical counterpart, but I can already assure the reader that my acquaintance based PrelDRSs for de re/de se belief reports will be fully compositional, as I represent both the embedded subject representation and the acquaintance relation in situ—at least, at the level of PrelDRS construction. Of course, this is a rather superficial kind of compositionality, in a sense comparable to the static trick of the free variable discussed in 2.3.2.1, in that the final output of the interpretation is by no means derived compositionally. Recall from 3.2.2.1 that that was never our intention: DRT+PA’s two-stage architecture distinguishes a compositional PrelDRS construction mechanism, and a pragmasemantic resolution algorithm that must be systematic and computable but that can never be fully compositional in the strict Montagovian sense in which we have been using that term. In short, PrelDRSs are generated compositionally and that’s all that matters.
3.3 Layered DRT

The main aim here is to provide DRT with a sound, two-dimensional account of semantic rigidity, and use that to counter the obvious objection that treating proper names and indexicals as presuppositions amounts to descriptivism, and is therefore easily rebutted by the so-called Kripke-Kaplan (KK) argumentation. It so happens that the very general framework developed by Geurts & Maier (2003), called L(ayered )DRT, is well suited to deal with this task. In essence LDRT is a simple extension of DRT that allows the information conveyed by an utterance to be stored in different layers of the same DRS. For instance, there could be a layer for representing the ordinary truth-conditional contribution, one for implicatures, and one for merely formal information (like the gender features on nouns in Germanic languages). As such, LDRT fits in with the recently popular family of multidimensional semantic frameworks, somewhat related to, for instance Potts (2003) or Dekker (2002), but more general. Let’s focus on the interaction between assertion and implicature a bit.

(77) There was a nice philosopher

\[
\begin{array}{c|c|c}
  & x & \\
\hline
\phi & \text{philosopher}(x) & \\
\text{n} & \text{nice}(x) & \\
\end{array}
\]

This DRS represents the truth conditional, or assertoric, information conveyed by an utterance of the sentence, but the choice of words in the predicate suggests that that philosopher is not the coolest person you’ve ever met, though nice doesn’t really exclude that interpretation, since the set of very cool individuals is a subset of the set of nice ones. This inference may be derived by Gricean pragmatic principles as follows: the term nice evokes a lexical scale of non-negative evaluative predicates, each more exclusive than the previous, e.g., nice < cool < fantastic. Obviously, the higher up in the scale, the stronger/more informative the predication would be. If we then assume that a speaker seeks to maximize informativity, we get that his use of the term nice implies that that philosopher was not really great, or he would have used a stronger, more appropriate term. The information so derived however is not of the same status as the rest of the information conveyed by the sentence. An important feature of these so-called scalar implicatures is that they are easily canceled by contradicting information. Consider the

\[33\text{A nice and comprehensive overview of the various forms of multidimensional semantics is Potts’s (2004b).}\]
following continuation of (77):

(78)  There was a nice philosopher
      In fact, she was great

Obviously the implicature that the philosopher is not great is canceled by this continuation. Note that it’s almost impossible to cancel assertoric information in the same way:

(79)  There was a nice philosopher
      a. ??In fact, she was not nice at all
      b. ??In fact, she was not a philosopher

The upshot of this is that, if we want to represent implicatures as conditions in a DRS, we must take care to mark their distinct, defeasible status. Well, that’s easy enough, take for instance Levinson’s (2000) idea of coloring DRS conditions, blue for assertion, red for implicature. In LDRT we use labeling, e.g. \(fr\) for assertoric content, \(imp\) for implicature:

\[
\begin{array}{c|c|c}
& x_{fr} & \\
\hline
\text{philosopher}_{fr}(x) & \text{nice}_{fr}(x) & \neg_{imp} \text{great}_{imp}(x) \\
\end{array}
\]

(80)  \[\triangle\\

The objection is that so far that’s just syntax, we need a way to cash out these colors semantically if it is to mean anything. And that’s where LDRT proves superior to Levinson’s coloring: it not only slices the DRS into different layers, but also adjusts the Tarskian truth definition to be sensitive to the layered composition. Details follow in 3.3.1, but the idea is that for any given combination of layers, we can compute the semantic contribution of those layers. For instance we can compute the Fregean contribution of (80), denoted \(\llbracket (80) \rrbracket_{fr}\), or the full pragmasemantic contribution, \(\llbracket (80) \rrbracket_{fr,imp}\), by simply ignoring conditions that fall outside the specified set of layers.

\[\text{34Unless by means of a denial, which is a kind of speech act usually restricted to a dialogue where one speaker corrects another. However, denials can occasionally occur in monologue, in cases where a speaker realizes he has made a mistake and corrects himself, as in:}\]

(i)  There was a nice philosopher. No wait, she was probably a linguist, come to think of it.
3.3 Layered DRT

Let’s briefly consider some applications of the LDRT framework. Geurts & Maier (2003) focus mainly on solving so-called ‘binding problems’ which have been around in presupposition theory for a long time (Karttunen & Peters 1979), but which arise for implicature as well. Take (77); on a simple theory of implicature representation, one that computes and represents implicatures at the sentential level and separately from other content, this statement would assert that there was a nice philosopher, and implicate that there was a not so great philosopher. Obviously that’s too weak, the lexical scalar implicature evoked by the use of nice should really be bound in the Fregean layer, i.e. there was a nice philosopher, and that philosopher was not great. In LDRT we get this by allowing fr-labeled discourse referents to bind imp-labeled atomic conditions. In other words, discourse referents are the glue connecting the different layers.

A proper representation and segregation of semantically interpretable types of information other than plain old Fregean truth conditions comes in handy when analyzing denial, which can be targeted at different layers resulting in e.g. Fregean (81a), implicature (81b), or presupposition (81c) denials (Horn 1985; van der Sandt & Maier 2003):

(81)

a. That’s not a donkey; that’s a mule
b. She didn’t get sufficient marks; she got all A’s
c. He didn’t stop smoking; he never smoked

To unify these, Maier & der Sandt (2003) propose a mechanism that removes ‘offensive content’ from a DRS representation of a dialogue to restore full cross-layer consistency. It is therefore crucial that we represent not just Fregean content, but also implicatures and (accommodated) presuppositions, and whatever other content is correctable in this way, and that we can interpret all these layers of content semantically in order to determine what causes inconsistency.

The third application of LDRT, the one that’s most interesting for us here, has to do with rigidity and indexicals. The idea, developed to some extent by Geurts & Maier (2003), is to introduce a layer, labeled k(ripke/kaplan), for the content of proper names and indexicals. Complicating the semantics a little, we can ensure that the content of these expressions is evaluated at a semantic context parameter. We’ll see in 3.3.2 how this effectively solves the Kripke/Kaplan objections against descriptivism. In 3.3.3 I’ll combine it with the presuppositional account of referential terms to do justice to some subtle data involving apparently non-rigid uses of proper names.
3.3.1 Basic LDRT

3.3.1.1 Syntax

In this section I give the syntax and semantics of basic LDRT. First the syntax, which is real simple. We extend our primitive inventory of discourse referents, predicates, modal operators and quantifiers with a finite set of layer labels, say \( \mathcal{L} = \{ \text{fr}, \text{imp}, \text{form}, k, \ldots \} \). If we attach one of these labels to a discourse referent we get a labeled discourse referent: \( x_{\text{imp}} \), \( z_{\text{form}} \). Attaching them to atomic conditions gives atomic labeled conditions: \( x =_{\text{fr}} y \), \( \text{donkey}_{k}(x) \). An LDRS simply pairs a set of labeled discourse referents with a set of labeled conditions, which is either atomic, or something like \( \neg_{\text{imp}} \varphi \), \( \Box_{k} \varphi \), or a duplex condition with a labeled quantifier such as MOST_{k}.

Let’s illustrate the basic syntax of the LDRS language with an example:

\[(82)\] Some of my friends own a donkey

\[
\begin{array}{c}
\text{x}_{k} \\
\text{speaker}_{k}(x) \\
\text{y}_{\text{fr}} \\
\text{friend}_{\text{fr}}(y, x) \\
\text{z}_{\text{fr}} \\
\text{donkey}_{\text{fr}}(z) \\
\text{own}_{\text{fr}}(y, z) \\
\end{array}
\]

In the following, we often avoid all too tedious labeling by the notational convention that everything in a subDRS belongs to the layer of the condition in which it is embedded, unless explicitly labeled otherwise. For instance, in (82), one \( \text{imp} \) label, on the negation, would be enough. We may further abbreviate an identical condition occurring at more than one layer at once, by subscripting it with multiple labels, i.e. \( \text{walk}_{k, \text{form}}(x) = \text{walk}_{k}(x) \text{walk}_{\text{form}}(x) \).

Whether a discourse referent should be considered bound or free depends on what layers you’re looking at.\(^{35}\) In (82), for instance, \( x \) is free in the

\(^{35}\)I’m currently looking into the possibility of foregoing the labeling of discourse referents. One advantage is that it would simplify the syntax and semantics somewhat, because many more DRSs would count as closed so semantic values are more often defined. Also, it would fit well with the idea of discourse referents as cross-layer ‘pegs’ that serve merely to connect layered information. On the other hand we do lose information that seems
3.3 Layered DRT

fr and imp layers, but bound if we take the k into consideration. This has repercussions for the notions of free variable, accessibility and proper embedding. First an auxiliary definition:

(83) For \( L \subseteq \mathcal{L} \), and \( \varphi = \langle U(\varphi), C(\varphi) \rangle \) an LDRS:
- a. \( U_L(\varphi) = \{ \xi \mid \text{there is an } l \in L \text{ with } \xi_l \in U(\varphi) \} \)
- b. \( C_L(\varphi) = \{ \gamma \in C(\varphi) \mid \gamma \text{ bears a label } l \in L \} \)

(84) \( FV_L(\varphi) = (\bigcup_{\gamma \in C_L(\varphi)} FV_L(\gamma)) \setminus U_L(\varphi) \)

(85) If \( \gamma \)'s label, \( l \), is not contained in \( L \) then \( FV_L(\varphi) = \emptyset \), otherwise, one of the following applies:
- a. \( FV_L(\Pi_l(\xi_1, \ldots, \xi_n)) = \{ \xi_1, \ldots, \xi_n \} \)
- b. \( FV_L(\neg \varphi) = FV_L(\varphi) \)
- c. \( FV_L(\varphi \lor \psi) = FV_L(\varphi) \cup FV_L(\psi) \)
- d. \( FV_L(\text{Bel}_l \varphi) = \xi \cup FV_L(\varphi) \)
- e. \( FV_L(\varphi \Rightarrow_l \psi) = FV_L(\varphi \langle \text{most}_l \Xi \rangle \psi) \)

Applied to example (82), which consists of three layers, k, fr, and imp, this means, for instance: \( FV_{\{k,fr,imp\}}((82)) = FV_{\{k\}}((82)) = \emptyset \), and \( FV_{\{fr,imp\}}((82)) = \{x\} \).

(86) \( f \) is an \( L \)-proper embedding for an LDRS or labeled condition \( \varphi \)
iff \( \text{Dom}(f) \supseteq FV_L(\varphi) \)

For example: \( \emptyset \) is an \( L \)-proper embedding for (82) iff \( k \in L \).

3.3.1.2 Semantics

The semantics of this intensional, layered fragment then. As stated before, the idea is to make interpretation relative to a set of layers. We need not touch the definition of a model at all:

(87) \( \mathcal{M} = \langle D, W, \mathcal{I}, \text{Bel} \rangle \)

We must now define verification of LDRSs and conditions relative to set of labels and a possible world: \( \mathcal{M} \models^f_{w,L} \varphi \). One complication is the already crucial, esp. when considering layered presupposition resolution and cases where you’d want undefined semantic values. It remains to be seen if these hurdles can be overcome in the interest of simplicity.
noted layer dependence of $FV$: individual layers of an LDRS often contain free variables which makes them uninterpretable on their own, while a combination of layers might make the same LDRS closed. This necessitates a partial semantics with special attention for definedness:

(88) If $f$ is an $L$-proper embedding for $\varphi$:
\[ \mathcal{M} \models^f_{w,L} \varphi \text{ iff there is a } g \supseteq_{U_L(\varphi)} f \text{ and for all } \gamma \in C_L(\varphi): \]
\[ \mathcal{M} \models^g_{w,L} \gamma \]
Otherwise, undefined

(89) If $f$ is an $L$-proper embedding for $\gamma$:
\[ \mathcal{M} \models^f_{w,L} \gamma \text{ iff } \gamma^\prime\text{’s label } l \text{ does not appear in } L, \text{ or one of the following applies:} \]
(i) \[ \mathcal{M} \models^f_{w,L} \xi =_l \eta \text{ iff } f(\xi) = f(\eta) \]
(ii) \[ \mathcal{M} \models^f_{w,L} \Pi(\xi_1, \ldots, \xi_n) \text{ iff } \langle f(\xi_1), \ldots, f(\xi_n) \rangle \in I(\Pi) \]
(iii) \[ \mathcal{M} \models^f_{w,L} \neg \varphi \text{ iff there is no } g \supseteq_{U_L(\varphi)} f \text{ with } \mathcal{M} \models^g_{w,L} \varphi \]
(iv) \[ \mathcal{M} \models^f_{w,L} \varphi \Rightarrow \psi \text{ iff for all } g \supseteq_{U_L(\varphi)} f \text{ with } \mathcal{M} \models^g_{w,L} \varphi \text{ there exists a further extension } h \supseteq_{U_L(\psi)} g \text{ with } \mathcal{M} \models^h_{w,L} \psi \]
(v) \[ \mathcal{M} \models^f_{w,L} \varphi \langle \text{MOST}_l \Xi \rangle \psi \text{ iff for most extensions } g \supseteq_{U_L(\varphi) \Xi} f \text{ with } \mathcal{M} \models^g_{w,L} \varphi \text{ there exists a further extension } h \supseteq_{U_L(\psi) \Xi} g \text{ with } \mathcal{M} \models^h_{w,L} \varphi \odot \psi \]
(vi) \[ \mathcal{M} \models^f_{w,L} \text{BE}l_{\xi_l} \varphi \text{ iff for all } v \in Bel(f(\xi), w): \mathcal{M} \models^f_{v,L} \varphi \]
Otherwise, undefined

Extensional and intensional semantic values (of LDRSs and conditions) are defined as follows:

(90) a. \[ [\varphi]_{w,L}^f = 1 \text{ iff } \mathcal{M} \models^f_w \varphi, \text{ and } 0 \text{ if } f \text{ is } L\text{-proper for } \varphi \text{ but } \mathcal{M} \not\models^f_w \varphi \]
b. \[ [\varphi]_L = \left\{ w \in W \mid [\varphi]_{w,L}^f = 1 \right\} \]
c. \[ [\varphi]_L = [\varphi]_L^0 \]

Just to be sure, let’s interpret some layers of our earlier example (82):

(91) a. \[ \{(82)\}^{k,fr,imp} = \text{ the set of worlds where there is a speaker, } x, \text{ and some but not all of } x\text{’s friends own a donkey} \]
b. \[ \{(82)\}^{k,fr} = \text{ the set of worlds in which there is a speaker, } x, \text{ and at least one of } x\text{’s friends owns a donkey} \]
c. \[ \{(82)\}^{fr} \text{ undefined, since } x \in FV_{\{fr\}}((82)) \]
d. \[ \{(82)\}^{x-Emar} \text{ = the set of worlds in which at least one of my friends owns a donkey} \]
In (91b) we interpret the indexical as a description, so the real asserted content must be something else. In fact, the result in (91d) looks quite right, but how exactly did we determine that partial embedding to close off the open layer encountered in (91c)? The next section is dedicated to a two-dimensional treatment of $k$ that will answer this.

### 3.3.2 Direct reference vs. descriptivism in LDRT

I’ll start by applying the crucial Kripke-Kaplan argument against descriptivism to DRT to reaffirm the need for rigidity. Then we’ll see Geurts’s (1997) defense of an PA-enhanced descriptivism of directly referential terms. Unfortunately, this framework still fails the KK test. Next, we discuss a different approach to rigidity in DRT, using ‘external anchors’, which we criticize for being, in a sense, too rigid. Then, I propose my own layered alternative and show how it passes the KK test. The really interesting data will have to wait until we can combine the layered account with the presuppositional one in 3.3.3.

### 3.3.2.1 Against descriptivism

Following Russell (1905), various authors have tried to analyze proper names as hidden descriptions specifying a list of properties that refer to a unique individual (Kneale 1962; Dummett 1973; Katz 1977, 2001). The standard example was the definition of ‘Aristotle’ as ‘the famous philosopher who was a student of Plato, teacher of Alexander the Great, born in Stagira, 384BC, died in Chalcis, 322BC,

...’ This was supposed to explain the fact that ‘Aristotle’ refers to Aristotle, by reducing the semantic contribution of the proper name to that of a definite description, often understood as quantificational, via Russell’s Theory of Descriptions. The main problem with this type of account is that if ‘Aristotle’ were really to abbreviate a full description of Aristotle, we’d predict that Aristotle was a student of Plato and he died in 322BC expresses a tautology. This is obviously wrong: Aristotle might well have had a different teacher, and he might have contracted some kind of disease and died five years earlier.\(^\text{36}\)

As Kripke (1972) notes, this argument is applicable to any kind of descriptive analysis of names, so we can’t get out of it by choosing only universally known properties of Aristotle to define the meaning of Aristotle. Take for instance the most uncontroversial descriptive definition we can think of:

\(^{36}\)Formulated in more metaphysical terms, none of the characteristics listed seem to be essential properties of Aristotle, in the sense that without any one of them he would still be Aristotle.
Chapter 3. A semantics of attitude reports

Aristotle = the person called Aristotle. Surely everybody will agree with that, it’s as close as you’ll get to a tautological (and readily generalizable so semantically useful) reduction of the name to a description of its bearer. But if we test substitutability in intensional contexts it’s easily seen that even this minimal descriptive analysis fails miserably:

(92) a. Aristotle is Aristotle
    b. Aristotle is [the person] called Aristotle
    c. It’s necessarily true that Aristotle is Aristotle
    d. #It’s necessarily true that Aristotle is called Aristotle
    e. #Some historians now claim that Aristotle was not really Aristotle
    f. Some historians now claim that Aristotle was not really called Aristotle

These examples strongly suggest that being called Aristotle is but an accidental property of Aristotle, and thus the distribution of a proper name X is not captured by the descriptive paraphrase the person called X.

In (92), intensional contexts were needed to bring out this distributional difference, but in fact, we don’t really need embedding at all. Let’s take a closer look at the unembedded equality (92a) and its descriptive counterpart (92b). The first is a real truism of the form $x = x$, it is necessarily true and as uninformative as it gets. But the second, though obviously true and highly uninformative, does not express a necessary truth. In terms of possible worlds, $[(92a)] = W$ (the tautological proposition), while $[(129)] = \{\text{worlds where Aristotle is called Aristotle, i.e. excluding worlds where his mother changed her mind and called him Thrasyboulos.}\}$

This sometimes forgotten contrast in unembedded sentences nips in the bud an otherwise attractive way out for the descriptivist. This attempted fix uses scope to provide an explanation of the observed contrasts between proper names and their descriptive paraphrases. Let’s take in some predicate logic with $\iota$-terms and for every proper name a (non-rigid) predicate meaning ‘the person called such-and-such’, i.e. $[\text{aristl}]_w$ = the set of people called Aristotle in $w$. With this, we can represent both narrow (93a) and wide scope (93b) readings:

(93) #It’s necessarily true that Aristotle is called Aristotle $[= (92d)]$
    a. $\supset \Box[\text{aristl}(\iota x. \text{aristl}(x))]$
    b. $\supset \# \exists y [y = \iota x. \text{aristl}(x) \land \Box[\text{aristl}(y)]]$

\footnote{I prefer to use . . . is called . . . instead of . . . is the person called . . . for readability. If not strictly equivalent in cases of multiple bearers of the same name, the first certainly is a logical consequence of the second.}
3.3 Layered DRT

Apparently, the false wide scope reading is preferred over the trivial narrow scope one. By postulating widest possible scope for our descriptions we can get the right truth conditions for this and other embedded proper name examples, like (92c) and (92f). However, if there’s no operator, there’s no scope, so for unembedded examples the wide scope stipulation doesn’t help at all:

\[(94)\]

\[
\begin{align*}
\text{a. Aristotle is Aristotle} & \quad [=(92a)] \\
\quad \triangleright \quad \forall x.\text{aristl}(x) = \forall y.\text{aristl}(y) \\
\text{b. Aristotle is called Aristotle} & \quad [=(92b)] \\
\quad \triangleright \quad \text{aristl}(\forall x.\text{aristl}(x))
\end{align*}
\]

It’s easy to see that these lfs are equivalent, \[[(94a)] = [(94b)] = W\], so we have no account of the intuitive difference in metaphysical status—the contingent status of (94b) in particular—that we observed above. So let’s leave descriptivism for now and look at the alternative.

Kripke’s solution is that proper names are not abbreviations or hidden descriptions, they are simply tags that are ‘attached’ to an individual by an act of ‘baptism’. Once fixed and in use, this arbitrary connection between the name and its bearer makes continued referential use of the name possible. In some cases, such uses are transmitted through the ages, as with Aristotle, resulting in so-called causal-historical chains from the current use to the initial baptism. Philosophical and historical subtleties aside, what matters for us is that names have no hidden descriptive content that enables users to find the referent, they just refer, just like the filename ch3_semantics.tex refers to a particular text file I’ve created a little while ago.38 I feel such a filename example brings out especially clearly what is meant by Kaplan’s (1989:483) saying that proper names and indexicals “refer directly, without the mediation of a Fregean Sinn as meaning”, which sums up the theory of direct reference.

Technically, Kripke and Kaplan work this out by assigning rigid intensions to proper names, to capture the fact that they really have no intensional content, just their referent, while staying within the intensional model theoretical framework. Applied to predicate logic this means we can represent proper names by individual constants, but assign these constants a rigid, i.e. world-independent, denotation (cf. 1.2.2.1).39 A definite (linguistic) advant-

38 Geurts (1997) uses a filename example to question the thesis that the proper referential use of names is always grounded in a social network.

39 A conceptually more direct formalization of direct reference would be the structured propositions approach in which the proper name’s referent can literally be contained. We’ve discussed this already in 1.2.2.1 and 2.3.3.1.
tage of doing things this way is that proper names are simply represented in situ: no scope, no movement, no problems with compositionality.

\[(95)\]

\begin{enumerate}
  \item It’s necessarily true that Aristotle is Aristotle \(\models \Box [a = a]\)
  \item \#It’s necessarily true that Aristotle is called Aristotle \(\models \Box [\text{aristl}(a)]\)
\end{enumerate}

If we assume that \(a\) rigidly refers to the actual Aristotle, in every possible world, the truth conditions of these simple lfs are exactly as we want them:

\[\llbracket \Box [a = a] \rrbracket_w = 1\] regardless of \(w\), but \(\llbracket \Box [\text{aristl}(a)] \rrbracket_w = 0\) because there are \(w'\) where \(\llbracket a \rrbracket_{w'} = \llbracket (a) \rrbracket \not\in \llbracket \text{aristl} \rrbracket_{w'}\) (i.e., worlds where Aristotle got a different name).

Because of the strong compositionality of the representations in (95), the truth conditions reflect a contrast in the underlying simple sentences:

\[(96)\]

\begin{enumerate}
  \item Aristotle is Aristotle \(\models (92a)\)
    \item Aristotle is called Aristotle \(\models (92b)\)
\end{enumerate}

By the same reasoning as above: \(\llbracket a = a \rrbracket = W \neq \llbracket \text{aristl}(a) \rrbracket\). To sum up, the view that proper names are directly referential is both theoretically and empirically superior to descriptivist attempts, with or without a wide scope assumption.

Most of the above applies to indexicals as well as proper names. If we’d try to reduce \(I\) to a description, the most likely candidate would be the speaker. Now consider some substitution instances similar to the ones in (92):

\[(97)\]

\begin{enumerate}
  \item I am speaking
  \item The speaker is speaking
  \item \#I’m necessarily speaking
  \item ?The speaker is necessarily speaking
\end{enumerate}

The judgments are relatively clear: an utterance of (97a) is always true and therefore uninformative, but it is never necessarily true, it does not express a tautology. This is corroborated by the fact that (97c) is so plainly false, because it’s easy to imagine an alternative state of affairs where I realized that silence is golden and not say anything at that particular point in time. On the other hand, (97b) does express a tautology, and (97d) is ambiguous between a false (‘wide scope’) and a tautological (‘narrow scope’) reading. By assuming \(I\) corresponds to a widest scope description the speaker we can get the embedded cases right, but the necessity contrast between (97a) and
Kaplan (1989) proposes that indexicals, like proper names are directly referential: If I utter I, I rigidly/necessarily refer to me, Emar. The tricky thing is how to reconcile this rigidity with the obvious context dependence of indexicals—when you utter I it refers to you. Kaplan’s solution has been discussed in detail in 1.2.4.2, but summarizing: we need two dimensions of intensionality: one for context dependence, one for modal, intensional phenomena. The resulting semantics assigns every expression a character, which can be thought of as a function mapping contexts to intensions, which in turn, as usual, consist of functions from worlds to extensions. Intensional rigidity (which makes them immune to embeddings under modal operators) and context dependence are thus united in a two-dimensional framework in which they live orthogonally. Applied to (97) we make the right predictions, viz.: If c is some context centered around me, Emar, then the proposition expressed by uttering (97a) in c = [[(97a)]^c = the set of worlds where Emar, the contextually determined speaker, is speaking; [(97b)]^c = the set of worlds whose speaker is speaking = W; [(97c)]^c = the set of worlds where Emar is necessarily speaking = ∅; (97d) is scopally ambiguous, but there’s no indexicality involved there.

3.3.2.2 Descriptivism revisited: proper names as presuppositions

Kripke’s and Kaplan’s arguments had converted many semanticists to the direct reference paradigm, because its account of proper names and indexicals was evidently superior to its descriptivist rivals. However, the subtle contrast in unembedded sentences, (92a)-(92b) and (97a)-(97b), is sometimes forgotten, leading to the defense of wide scope descriptivism in some form or other. I will argue that Geurts’s (1997) account of proper names as presupposition commits this fallacy, as does the analogous account of indexicals by Hunter & Asher (2005). Showing this, I’m merely reformulating the argumentation of Zeevat (1999) who already wielded the old KK argument against (his own) presuppositionally enhanced descriptivist account of indexicals.

The basic idea of this neo-descriptivism is that names and indexicals are definites, just like definite descriptions and anaphoric pronouns, and as such they are presupposition triggers. Taking the minimal descriptive content as presupposed content we automatically derive the observation that their reference is consistently unaffected by modal operators from the projection behavior of presuppositions. Despite a more sensible formulation in dynamic semantics, this is essentially a wide scope descriptivist account, since the result of presupposition projection in these cases is always an output DRS equivalent to an If with wide scope descriptions.
(98) a. #It’s necessarily true that Aristotle is called Aristotle \[= (92d) \]

Note that the conditions used are still the same descriptive predicates as before: \( \llbracket \text{aristl} \rrbracket_w = \) the set of people called Aristotle in \( w \); \( \llbracket \text{speaker} \rrbracket_w = \) the set of speakers in \( w \). Note further that with the transition from static to dynamic, I’ve put an Aristotle and a speaker in the input contexts because otherwise there would be nothing to bind to (and the SPECIFICITY constraint might block accommodation). I agree with Geurts that proper names can occasionally be used to introduce a new discourse entity (through accommodation, cf. 3.3.3.2), but at least for \( I \) there is really no need for accommodation, since every proper speech context surely contains a highly salient speaker.

It’s now easy to see that these outputs represent the correct truth conditions. Unfortunately, replacing the proper name or indexical with its de-
scriptive counterpart gives us the exact same preliminary DRS and outputs, so we’re at a loss to explain the observed contrast between, say, (97c) and (97d). Perhaps we could try to fix this, for instance by stipulating that the presupposition triggered by I is different from that triggered by a surface realized definite description the speaker in that the former really has to float up to the main context, whereas the latter is more easily bound or even accommodated locally. Hunter & Asher (2005) propose a fix along these lines, but before discussing the pros and cons of such a stipulation, there’s another problem to worry about: the unembedded cases. Again, if there’s no operator, there’s only one place for the presupposition to go, so there can be no distinction between name/indexical presuppositions and those triggered by the corresponding genuine definite descriptions. Crucially, for the proper name case:

\[
\begin{align*}
\text{(99) a. Aristotle is Aristotle} & \quad [= (92a)] \\
\end{align*}
\]

\[
\begin{array}{l}
\text{x = y} \\
\text{x} \\
\text{aristl(x)} \sim \text{aristl(y)} \\
\hline
\end{array}
\]

\[
\begin{array}{l}
\text{x} \\
\text{aristl(x)} \\
\end{array}
\]

\[
\begin{array}{l}
\text{x} \\
\text{aristl(x)} \\
\end{array}
\]

\[
\begin{array}{l}
\text{x} \\
\text{aristl(x)} \\
\end{array}
\]

\[
\begin{array}{l}
\text{x} \\
\text{aristl(x)} \\
\end{array}
\]

Both outputs of course are equivalent, moreover they are equivalent to the input contexts. In any case, these DRSs in no way reflect the different necessity status of (92a) and (92b) that played such a crucial role in the KK argumentation. At this point I think we can discard the presuppositional theory of indexicals and proper names as it is presented above—which is not to say that Geurts doesn’t have some important criticism for rigid direct
referentialist accounts, which we shall examine more closely in 3.3.3 after we have first introduced such an account of rigidity in DRT.

3.3.2.3 Real rigidity in DRT: external anchors

The Kamp & Reyle (1993) textbook apparently does take the KK arguments seriously and consequently tries to build in real rigidity. The idea is that apart from constructing a mental representation, the purely symbolic DRS, a hearer connects certain discourse referents to their actual referents in the outside world, if available and if so prompted by the speaker’s words. This last condition is to properly tie rigidity to certain linguistic items, most notably to proper names and indexicals.

Technically, these links, called external anchors, are simply restrictions on truthful embeddings. For ease of discussion, let’s illustrate Kamp’s external anchoring with the proper name example:

(100) Aristotle is called Aristotle

\[
\begin{array}{|c|}
\hline
x \\
\hline
\end{array}
\sim
\begin{array}{|c|}
\hline
\text{aristl}(x) \\
\hline
\end{array}
\begin{array}{|c|}
\hline
x \rightarrow \text{Aristotle} \\
\hline
\end{array}
\]

On the left, the DRS; on the right, the anchor, a partial function from discourse referents to individuals. Let’s not go in to the exact two-stage derivation of this anchored DRS yet, but concentrate on the endpoint and its interpretation. We see that the predication is represented as a DRS condition, as usual, but the proper name subject is lifted out of the DRS into the anchor, leaving but a bare discourse referent. The anchor, interpreted as a restriction on embeddings, thus only contributes referents on a metalevel, so to say, the information contained therein is in no way subject to operators or any kind of hierarchical DRS structure in general. This immediately apparent from the way (100)’s truth conditions are computed:

\[
[(100)]^f = [(100)]^{f \cup \{x \rightarrow \text{Aristotle}\}} = \left\{ w \mid \text{aristl}(x)_{\lfloor w \rfloor} = 1 \right\}
\]

In short, we get the correct, KK-proof proposition about the actual Aristotle.

The Kampian story behind the construction of (100) is as follows: the hearer noticed that the speaker used a proper name with referential intentions and thus searched his surroundings in the actual external world for an individual that would match these intentions, in this case a (salient) individual called Aristotle. He found Aristotle and extended his anchor accordingly, i.e.
3.3 Layered DRT

with a link mapping the subject term’s discourse referent to this Aristotle.

There’s a couple of things to consider here, all of them combined cast-
ing serious doubt on the appropriateness of external anchoring for treating
rigidity in DRT. First and foremost, the theoretical status of anchors is a
bit unclear: on the one hand they are part of the interpretation process,
they are constructed in parallel with the DRSs and share discourse referents
with them, but on the other hand anchors are really partial embeddings,
which live in the meta language and contain actual individuals taken from
the model’s domain. It seems they are constructed and represented alongside
DRSs but do not belong to the same object language. Such confusion would
be problematic for classical theories, but for DRT it is even more worrisome,
for parallel to the object-/metalanguage distinction there’s the narrow/wide
one that’s especially relevant for DRT, which was originally intended as a
purely symbolic, representational theory of interpretation as part of human
cognition. Remarkably, it was Kamp himself, an outspoken proponent of the
mentalistic conception of DRT, who introduced a mechanism to encode a
definitely wide aspect of semantics. It is this methodological inconsistency
that led others, like Geurts and van der Sandt (p.c.), to drop anchors in favor
of the unified account of referential terms as presuppositions, even if their
conception of DRT was less obviously mentalistic. Kamp and Asher take a
different course, keeping the wide, external anchors, but supplementing them
with narrow, descriptive counterparts called internal anchors.

Second, I’m not quite sure how anchors fit in to the two-stage procedure of
DRT, i.e. are they constructed in the resolution process as just another type of
context dependency that needs to be resolved? Or is there a separate module
for constructing external links? More specifically, if we look at DRT+PA,
we might ask how does this relate to the basic idea that all definites are
treated as presupposition triggers? Does that not hold for directly referential
ones, or is the construction of anchors something that comes on top of the
presupposition resolution of proper names.

A third point concerns the way we determine what goes into an anchor.
The idea expressed in the ‘Kampian story’ above are clear enough, but don’t
we really want to see some of that hardwired into the semantics, rather
than all in metatheoretical talk? It seems a shame not to represent the
linguistically specified descriptive content of the proper name that is used
to retrieve the appropriate referent. The problem becomes more apparent
with indexicals: an utterance of I always refers to its speaker, that’s the
lexical meaning. But we’d lose that on a simple anchoring approach, because

\(^{40}\)We should mention here also Zeevat’s (1999) related approach of replacing external
anchors with so-called intensional anchors.
Chapter 3. A semantics of attitude reports

we’d first determine who said it and then simply put that individual into the anchor:

(102) #I’m necessarily speaking

\[
\begin{array}{c}
\blacktriangleleft \sim \# \\
\Box \text{speak}(x) \quad [x \mapsto Emar]
\end{array}
\]

Compare this to the step from Kripke’s theory to Kaplan’s two-dimensional semantics. In Kripkean semantics, as in anchored DRT, the way in which rigid designators get their reference is purely metasemantic, in Kaplan’s theory we get an extra level of semantic representation, the character, in which the descriptive content is encoded, at least for indexicals like \textit{I}.

Though it can be seen as a minimal repackaging of the anchoring semantics, with a slightly higher dose of Kaplanian character theory, my account of rigidity in LDRT eventually addresses all the above criticisms and then some, viz. the Geurtsian critique of direct reference from a dynamic standpoint, but part of it will have to wait until 3.3.3.

3.3.2.4 A two-dimensional semantics of the \textit{k} layer

We’ve seen two attempts at bringing some measure of rigidity for proper names and indexicals into DRT: wide scope (neo-)descriptivism and anchors. Each turned out unsatisfactory (for different reasons, both empirical and theoretical). I will currently show how LDRT can resolve the conflict between descriptive and anchoring approaches.

The first step is to represent both intensional and rigid content uniformly as descriptive conditions in a single LDRS, but at different layers:

(103) I am speaking

\[
\begin{array}{c}
\blacktriangleleft \sim \\
x_k \\
\text{speaker}_k(x) \\
\text{speak}_k(x)
\end{array}
\]

On the surface, this already gives us a tangible difference between \textit{I} and the spelled out description \textit{the speaker}, which was exactly what was missing from the descriptivist accounts: \textit{I} and \textit{the speaker} each contribute the same discourse referent+condition, but \textit{I} puts it in the \textit{k} layer, while the overt description may put it in the \textit{fr} layer.\footnote{As we’ve noted before (p. 49), definite descriptions may also be used referentially, in}

\textit{the speaker}, which was exactly what was missing from the descriptivist accounts: \textit{I} and \textit{the speaker} each contribute the same discourse referent+condition, but \textit{I} puts it in the \textit{k} layer, while the overt description may put it in the \textit{fr} layer.\footnote{As we’ve noted before (p. 49), definite descriptions may also be used referentially, in}
3.3 Layered DRT

(104) The speaker is speaking

<table>
<thead>
<tr>
<th>x_f</th>
<th>~</th>
<th>speaker_f(x)</th>
<th>speak_f(x)</th>
</tr>
</thead>
</table>

But that’s just syntactic color, we need to show that we get different truth conditions for (103) and (104), and moreover, the right ones. Just applying the general LDRT semantics does not get us very far: \([104]_f = [103]_f,k = W\), but crucially \([103]_f\) is undefined. To check the KK judgments, we would like to compare the truth conditional, asserted contributions of the two, i.e. the interpretations of their \(f_r\) layer, but in (103) that layer is open. It seems we’re going to need a more sophisticated way to compute the Fregean contribution of an LDRS. We’ll define a notion of \(f_r\) content against a \(k\) background. The idea is we first compute the \(k\) layer’s contribution and then use that to create an anchor, against which we’ll be able to interpret the \(f_r\) layer. By using anchors we will have incorporated the strong, classical notion of rigidity, while still representing the lexical meaning of \(I\) in the LDRS with a descriptive condition.

To get all this right however we need to go two-dimensional, for if we want the \(k\) conditions to be interpreted, we have to make sure they’re interpreted with respect to a semantic context parameter distinct from the modal evaluation world. So we add a domain of contexts to our models, this time as a primitive, rather than using the traditional definition as triples, just for convenience:

(105) \(M = \langle D, W, C, I \rangle\)

What distinguishes contexts and possible worlds is that contexts always have a unique speaker, or rather ‘agent’ or ‘center’, because we will also consider contexts of thought. In other words: For every context \(c \in C\): \([\text{speaker}]_c\) is a singleton. And of course, just as in the standard formulation of Kaplanian semantics, each context comes with a unique world, time, and place as well. We sometimes use the familiar functions to pick out those coordinates: \(w_c\) is the world of \(c\), \(t_c\) the time of \(c\), etc.

Next, we turn to the truth definition. From now on, let’s concentrate on just the \(f_r\) and \(k\) layers, for after that, generalizations to arbitrary layer

which case they behave like rigid designators. On the current proposal we can handle this so-called referential/attributive distinction quite nicely by saying that definite descriptions can put their content either on the \(f_r\) or on the \(k\) layer of the DRS under construction, the choice driven by global pragmatic constraints, perhaps. The difference with indexicals is that those must always take the \(k\) layer.
sets, if necessary, can be generated easily. We’ve said that LDRS conditions always live on but a single layer, a complex label really abbreviates a number of copies of the condition on different layers. Here, this pays off, for it means a condition or discourse referent either carries a fr label, in which case it’s interpreted with respect to a \( w \in W \), or it’s in \( k \) and then it will be evaluated in a \( c \in C \): we don’t need to doubly relativize the interpretation of LDRS conditions to \( w \) and \( c \) simultaneously. I refer the reader who’s worried this constitutes an oversimplification to Zimmermann (1991; 2004). In these papers, Zimmermann defends his (1991:164) “Hypothesis (L): lexical items are always deictic or absolute”, i.e. an expression’s reference either depends only on the context (and is intensionally rigid), or the expression has intensional content but is contextually inert. The validity of this hypothesis would obviously constitute good evidence for our simplifying assumption about the fr/k distribution.

In any case, we need an LDRS condition’s truth in a model defined only with respect to a layer, an embedding, and a world or a context. The first option, \( \models_{w,fr} \gamma \), is already taken care of in (88)-(89); for the second, \( \models_{c,k} \gamma \), just replace all fr’s with \( k \)’s and all \( w \)’s with \( c \)’s. Now we want to combine these two into the desired content notion that turns \( k \) into an anchor. We use the following auxiliary definition:

\[
(106) \quad [\varphi]_{c,k}^f \text{ is the smallest } g \supseteq \Upsilon_k(\varphi) \text{ with } \mathcal{M} \models_{c,k}^g \varphi
\]

Note that \( [\varphi]_{c,k}^f \) is only defined if there is a unique truthful embedding of exactly the \( k \) universe into the context. If so, it outputs that embedding, which we can then use as anchor to close the fr layer rigidly, giving us the following notion of proposition expressed by \( \varphi \) in \( c \):

\[
(107) \quad [\varphi]_{f,c} = [\varphi]_{fr}^f [\varphi]_{c,k}^f
\]

The right side of the equation is defined in basic LDRT, cf. (90). It models the classical proposition expressed by the frege layer with the \( k \) labeled discourse referents rigidly fixed in \( c \). For mnemonic purposes, the idea behind my notation is that layers and worlds to be evaluated as propositional content are subscripted, while rigid, reference fixing materials are in superscript. For further notational brevity, an empty embedding may be omitted, as always.

To see that this \( [\varphi]^c \) really corresponds to the intuitive, classical notion of propositional content that we were after, let’s compute it for some of our earlier examples:

\[
(108) \quad \text{Let } c \text{ be a context where I utter the sentence in question}
\]
3.3 Layered DRT

These demonstrations conclude my reconstruction of a good classical content notion in LDRT. It wasn’t entirely trivial, and it may seem but a meagre justification for adopting LDRT, but we are really going to need this if we want to be able to even compare our results in report semantics with the results of rival classical theories in the broadly Kaplanian paradigm, such as Schlenker’s and von Stechow’s.

3.3.3 Presupposition and the two-stage architecture in LDRT

Until now we have restricted ourselves to representing and interpreting fully developed output contexts in LDRT. It’s time to worry about how to derive such representation from input contexts and (English) sentences. We want to stick with the Presupposition as Anaphora theory and its two-stage architecture. This means we have to think about layer-sensitive extensions of the Prel and Res algorithms. Geurts & Maier (2003) suggest we put presuppositions in their own layer, and in 3.3.3.1 we’ll flesh out the suggestion a little further to find out that it that cannot mean that triggered presuppositions in preliminary LDRS are represented at a presupposition layer. Instead, we keep the old $\partial$ operator/dashed box formalism for representing presuppositional material. In addition to this presupposition marking, LDRT still labels every condition and discourse referent and the combination of these independent markings has some very interesting consequences for the analysis of proper names and indexicals, which will be explored in 3.3.3.2. The analysis I propose there leads to a synthesis of neo-descriptivism and $k$-anchored rigidity that can handle both the KK judgments and Geurts’s (1997) tricky cases where these otherwise directly referential terms appear to be used non-referentially.
3.3.3.1 A layer for presuppositions?

Although Geurts & Maier (2003) are not too clear about it,\(^{42}\) they seem to suggest that we represent unresolved presuppositions by putting them in a pr layer, instead of using a $\partial$ operator, dashed box, or underlining to mark certain material in a PrelDRS ‘to be resolved’. For example:

\[ (109) \quad \text{The soup is warm} \quad \left[ \begin{array}{c} x_{pr} \\ \text{soup}_{pr}(x) \\ \text{warm}_{fr}(x) \\ \neg \text{imp}_{imp}(x) \end{array} \right] \]

The nice thing about this type of representation is the uniform treatment of presuppositions with say rigidly referential material and assertoric information. As a result, we can straightforwardly compute the presuppositional contribution of the sentence in abstraction of the discourse context, viz. $[(109)]_{pr} = \text{the proposition that there is soup}$. Additionally we predict that there is no purely Fregean proposition expressed, because the fr layer depends on pr, as does imp. It might be interesting to see what the anchoring trick used to turn $k$ into a background would yield if applied to pr, but not now.

On this view it’s reasonably clear what Prel has to do. Very roughly, it parses the sentence into constituents and categorizes them as presupposition triggers (e.g. definite descriptions), directly referential terms (e.g. indexicals), descriptive predications (e.g. verbs) and terms that evoke readily computable implicatures (e.g. scalar adjectives). Meanwhile it builds an LDRS in which all conditions and discourse referents bear the labels corresponding to these categories, e.g. fr, $k$, pr, or imp. The Res algorithm merges a PrelDRS with an input context, looks at everything new in the pr layer and tries to bind or accommodate it to suitable antecedents according to the preferences described in 3.2.2. Obviously this is nowhere near a complete picture of dynamic multilayered discourse interpretation; all layers considered so far have some special properties of their own, so each in isolation already requires extra work, not to mention the complicated interactions that occur between different layers, or other layers that will need to be distinguished. We’ve spent some time working on the semantics of the $k$ layer in 3.3.2.4, and now we’ll

\(^{42}\)To be honest, it looks suspiciously as if they are confusing two distinct notions of presupposition: utterance presupposition ($\Rightarrow$) and lexically triggered presuppositional material ($\partial$), cf. 3.2.1.
3.3 Layered DRT

take a close look at \textit{pr}, but take the \textit{imp} layer: a proper description of what goes on there must incorporate everything we know about the different kinds of implicatures and their computation and cancellation properties. Obviously I will not go into any of this; as it stands LDRT merely means to present a general, logical tool to make it possible to semantically \textit{represent} these different types of information and their interaction.

Although I agree with the very general story sketched above, I do not endorse the proposed representation of presuppositions (anymore). Let me mention two objections against the \textit{pr} layer before coming up with a better solution, viz. a return to van der Sandtian presupposition boxes. In fact, the strongest argument against \textit{pr} will be the successful application of the alternative to directly referential terms in the next subsection.

An immediate technical difficulty with \textit{pr} would be the representation of complex presuppositions for phrases like \textit{his donkey}. The \textit{pr}-style representation exemplified in (109) does not allow stacking of presuppositions in this way. The reason we’ve been using recursively embeddable boxes in the first place was to group together the material presupposed by a single trigger as a unit. With just a single \textit{pr} marked layer it will be difficult to properly distinguish distinct presuppositions triggered in a single sentence. And the other way around, a presupposition consisting of multiple conditions and a discourse referent may end up scattered throughout the DRS in resolution because \texttt{Res} has no way of knowing what parts of \textit{pr} belong together. Although Geurts (1999) gives some evidence that a presupposition’s discourse referent may sometimes end up higher than its conditions, I’m unconvinced that resolving all presupposed conditions and discourse referents completely independently (essentially putting a lot of faith in the \texttt{closure} constraint) will not overgenerate readings. And then there’s the lost feature of stacking van der Sandtian dashed boxes to form complex presuppositions, which was used primarily to mark the order of resolution for complex presuppositions. Take \textit{his donkey}; this triggers a complex presupposition in which the third person pronoun denoting the owner is to be resolved before we can start to think about the donkey. As it happens, in this simple case the resolution order is not so important, because the owned donkey presupposition logically depends on the pronominal element anyway, but it’s unclear whether complex presupposition embedding is always redundant in view of a global \texttt{ACCESS/CLOSURE} constraint.

The second objection concerns the supposed advantage of the \textit{pr}-layer approach, semantic interpretability of unresolved presuppositions. This turns out to be a red herring if we ask ourselves, do we really need to interpret PrelDRSs? And if so, are the proposed interpretations any good? As to the first question, I’d have to say, probably, yes, interpreting PrelDRSs seman-
tically would benefit the resolution process, for instance for checking consistency and informativity, as discussed in 3.2.2. The second question however must be answered negatively. What would be useful for Res is something like ‘the preliminary proposition expressed’ by a sentence, and we could be tempted to model this in LDRT as \[\ldbracket \{fr, pr\}\]}. Applied to the soup example: \[\ldbracket (109) \{fr, pr\}\] = the proposition that there’s warm soup. I guess that’s the best we can expect, apart perhaps from rigidifying the presupposed soup à la 3.3.2.4. If we look at PrelDRSs with embedded presuppositions this rigidification option no longer makes sense, and the combined \{fr, pr\} content just always represents the reading we’d otherwise get from locally accommodating the presupposition. But why would we get a useful notion of preliminary proposition by local accommodation of all presuppositions? That doesn’t even consistently correspond to either the weakest, or the strongest reading. Instead of considering all the options here I’ll leave this an open question, in order to save some space for my main argument against the pr layer, in favor of boxes.

But before we zoom in on directly referential terms again, I’ll first present my general alternative representation of van der Sandtian presuppositions in LDRT. In short: if \textbf{Prel} encounters a presupposition trigger, it creates an \(\partial\) marked LDRS with the presupposed material divided over a number of layers according to the type or status of the conditions and referents to be resolved. Proper names for instance trigger a presupposition with the name condition and discourse referent in the \(k\) layer, while definite descriptions will have presuppositions mostly in the \(fr\) layer. The presupposed discourse referent determines ‘the layer of the presupposition’, so proper names contribute so-called \(k\) presuppositions, and definite descriptions trigger \(fr\) presuppositions. The next logical step would seem to be the formulation of a binding constraint that says that a presupposition needs to be bound to something in its own layer. And that goes for accommodation as well: if binding fails, drop presupposed content in the presupposition’s layer, at a suitable accommodation site.

Let’s see some examples. Here’s one with a proper name and a definite description:

(110) Ligia doesn’t like the professor
3.3 Layered DRT

It’s true that this sentence can occur felicitously in a discourse context with a salient Ligia and a given professor. Take for instance a context where Ligia, a mutual acquaintance, is represented in the common ground’s $k$ layer, and a professor has just been brought up in the discussion, like so:

(111) [In Ligia’s group there’s a postdoc and a professor.]

Ligia doesn’t like the professor

Another example, this time with an indexical, triggered by a possessive construction, so embedded inside a descriptive presupposition (recall how presuppositional embedding was impossible with the $pr$ layer approach). After binding the first person possessive’s presupposition, we accommodate the descriptive presupposition, to illustrate also the layer-faithfulness constraint on accommodation.

(112) My best friend owns a donkey
Chapter 3. A semantics of attitude reports

These examples show the basic applications of the theory so far: a descriptive presupposition triggered in \( fr \) and then bound (111) or accommodated (112) in \( fr \), and a proper name (111) or indexical (112) triggering a \( k \) presupposition that gets bound in \( k \).

The ‘layered presuppositions’ analysis makes for a nicely compositional Prel, featuring a uniform treatment of all definites as presupposition triggers. On the other hand we avoid the KK criticism by having all proper names and indexical presuppositions end up in the \( k \) layer of the output, and then giving that layer a semantically rigid interpretation through the anchoring procedure of 3.3.2.4. But there’s more. By relaxing the layer-faithfulness constraint somewhat we can do justice to a host of alleged counterexamples to the classical theory of direct reference.

3.3.3.2 Cross-layer binding and non rigid proper names

Let’s take a closer look at layer-faithfulness and see if it really is empirically justified. One possibility excluded is that \( fr \)-presupposition are bound by \( k \)-antecedents. This means that once something is indexically identified in the common ground we can never refer to that entity with a description anymore. Note first that this seems to accord with the general observations in work on the givenness hierarchy (Gundel et al. 1993), also known as the familiarity (Prince 1981) or referential (Zeevat 2002) hierarchies.\(^{43}\) From these studies

\(^{43}\)This hierarchy of NPs amounts to something like: indexical pronoun > 3rd person pronoun > proper name > (complex) demonstrative > definite description > indefinite
there emerges a production preference ranking of NPs according to which indexicals and proper names are preferred over descriptive phrases. It seems to follow that if something is available in the $k$-layer, and thus available to be picked up by a demonstrative or higher, we can no longer use a definite description or indefinite to refer to it. This would corroborate our prediction that $fr$-presuppositions must be bound in $fr$.

However, there seem to be some counterexamples, like expressive descriptions, or epithets (Jackendoff 1972; Schlenker 2005), that are well-known for referring to entities evidently in the $k$-layer:

(113)  I wanted Charlie to help me, but the bastard wouldn’t do it.  

[[Jackendoff 1972:110]]

Here, Charlie is first referred to by a totally rigid proper name, and then later picked up by a definite description. Several theories have been put forth to account for the special status of epithets, some claiming that they are to be analyzed as a special type of pronouns, which leads us to the next problem for strict layer-faithfulness: anaphoric pronouns.

Obviously a pronoun like he in (114) can take both $fr$- and $k$-antecedents:

(114)  \[
\begin{cases}
  \text{He}_k \\
  \text{Colby}_k \\
  \text{That man}_k \\
  \text{The philosopher}_fr \\
  \text{A philosopher}_fr 
\end{cases}
\]  

beats every donkey he doesn’t like

The question of course is, what label can we put on a pronoun’s presupposition? Three options: none, both $fr$ and $k$, allow cross-layer binding. All three require substantial revisions to the basic LDRT framework. Cross-layer binding for instance requires that we rethink most of the layered binding constraints, e.g. what kind of layers do we consider in the computation of semantic values for checking MATCH or CONSISTENCY? And then then there’s accommodation to think about, but fortunately pronouns do not accommodate (SPECIFICITY)—at least not fully, sometimes a pronoun can add new information:

(115)  \[
\begin{cases}
  \text{A famous professor} \\
  \text{Prof. Rice}
\end{cases}
\]  

visited the department. She gave a talk.  

[$\approx$(57)]

This discourse was meant to show how a presupposition is partially bound, with some of its content accommodated to reveal the professor’s gender. In
the context of the layered presupposition debate, the first variation shows the pronoun bound in the \( \text{fr} \) layer dropping the accommodated part of its content at that same \( \text{fr} \) layer, while the genderless proper name variation arguably demonstrates a case of partial binding with accommodation into \( k \).

At this point the simplest solution would be to just leave the layer of pronominal presuppositions underspecified. We represent this by suppressing the labels. The \textbf{Res} algorithm tries to match and bind such unlayered presuppositions in each of the layers separately, eventually choosing the one that gives the best output.  

The result:

(116) A famous professor visited the department. She gave a talk.  

\[
\begin{array}{c|c|c}
\text{fr} & \text{y} & \text{fr} \\
\hline
\text{famous}_{\text{fr}}(x) & \text{prof}_{\text{fr}}(x) & \text{dept}_{\text{fr}}(y) \\
\text{visit}_{\text{fr}}(x,y) & \text{give}_{\text{fr}}(z) & \text{female}_{\text{fr}}(x) \\
\end{array}
\]

Since we’re on the subject of pronouns, let’s take a moment and decide what conditions to give them as content, because there is really more to \textit{she} than the fact that it must be bound by a female individual. In English, the morphologically realized features of gender, number and person are all semantically interpreted—in \( \text{fr} \) or \( k \), that is. In other words, these features all correspond to properties of \textit{individuals}; 3rd person, for instance, means that its (antecedent’s) reference must be distinct from both the speaker and her audience. We can contrast such \textit{semantic} features with \textit{formal} features, like (Romance) gender, which restricts resolution to referents derived from \textit{words} of the specified grammatical gender. Formal features would be stored in another special layer, \textit{form}, for the interpretation of which I refer the reader to Geurts & Maier (2003), because in the current work we will consider only semantic features in \( \text{fr} \) and \( k \). Here’s the representation and semantics I propose for pronouns (it works for all three persons and genders, though only in the singular; I leave the semantics of plurals for future research):

44Technically this is easily accomplished, for instance by copying the presupposition into the separate layers, resolving them independently and then checking which output is optimal.
3.3 Layered DRT

(117) a. **Prel:** $I \mapsto \{x \mapsto 1_{sg}k(x)\}$

$I(1_{sg})(w) =$ the set containing the center(s) (i.e. agent, speaker, or thinker) of the world (or context) $w$

b. **Prel:** $she \mapsto \{x \mapsto fem.3_{sg}k(x)\}$

$I(fem.3_{sg})(w) =$ the set of $d \in D$ which are female in $w$ and distinct from the center and the addressees (if any) of $w$.

I assume the underspecified layering is necessary only for third person, the other two are really indexical, so firmly in $k$. Or are they? At the end of the section we’ll come back to the resolution of the indexical person pronouns.

Now, there is a third set of data that seem to involve presuppositions changing layers, this time it’s $k$-presuppositions, triggered by proper names, that seem to end up in $fr$. The examples are below are mostly taken from Geurts (1997) who uses them to show that proper names exhibit all the behavior typical of presuppositions. First, there’s global binding, and sometimes the antecedent is descriptively introduced new material, so in $fr$ (the DRSs below represent the observed rather than the currently predicted resolution):

(118) I have a poodle named ‘Horace’. Horace is three years old.

[[Geurts 1997:321]]
Chapter 3. A semantics of attitude reports

The same kind of binding occurs non-globally:

(119) If a child is christened ‘Bambi’, then Disney will sue Bambi’s parents.

\[(\text{Geurts 1997:321})\]

A somewhat similar case of a proper name ending up in the restrictor’s \(fr\)-layer:

(120) Every time we do our Beatles act, Ringo gets drunk afterwards.

\[(\text{Geurts 1997:322})\]

The observed resolution of the proper name could be classified either as binding or as accommodation, but in any case it’s clear that \textit{Ringo} in the consequent does not refer to the actual Ringo Starr that may be available in
the main context.

Actually, there are also clear cases of proper names being accommodated in $fr$. For instance, if I want to introduce Colby to you, I might use a construction like:

(121) This is Colby

But the most blatant violation of the theory of direct reference cited by Geurts, is one where the proper name is accommodated locally in the $fr$-layer:

The electoral process is under attack, and it is proposed, in light of recent results, that alphabetical order would be a better method of selection than the present one. Someone supposes that ‘Aaron Aardvark’ might be the winning name and says, ‘If that procedure had been instituted, Ronald Reagan would still be doing TV commercials, and []

(122) [Aaron Aardvark might have been president

\[\text{(Bach 1987:146-7)}\]

This represents the intuitively optimal output: there might have been some guy called ‘Aaron Aardvark’ and then he might have become president. The utterance as a whole certainly does not commit the speaker to the existence of anyone named Aardvark, which in PA corresponds to a non-global resolution of the presupposition. In this case, that must mean local accommodation. Further, it’s evident that the proper name is used descriptively, i.e. the presupposition, triggered as $k$, ends up in $fr$, contributing someone called Aaron Aardvark rather than the actual person. Moreover, what would a $\Diamond_{fr}$-embedded $k$-layer mean anyway? If you think about it, there doesn’t seem to be any sensible way of interpreting material in an embedded $k$-layer.

The current proposal in effect ignores a $k$-layer embedded in a $fr$-labeled

\[\text{\footnote{Instead of, or in addition to, the $\Diamond$, we might just as well have chosen to embed the DRSs in a conditional’s consequent without it affecting the argument below.}}\]
Chapter 3. A semantics of attitude reports

condition in the computation of propositional content (i.e. $[\varphi]^c$, cf. (107)). This is in accordance with what we’re trying to capture with $k$: stuff at the main DRS can be interpreted as ‘contextually given’, but the $k$-labeled part thereof is contextual in a stronger, Kaplanian, truly semantic sense. However, if the analysis of attitude reports requires a context-manipulating belief operator, as argued in the previous chapters, we might find a use for embedded $k$-layers after all. We will take this up below and further in 3.3.4.

To sum up, proper names can be bound ((118), (119)) and accommodated ((121), (122)) in the $fr$-layer, though normally they end up in a $k$-layer in accordance with the theory of direct reference. Furthermore, these resolutions can be non-global ((119), (120) and (122)). This last fact would be good evidence for our view of proper names as presuppositions, if we could say something about the observed layer-switching to $fr$. And what to say about that really depends on how general the phenomenon is. In other words, do other $k$-presuppositions (i.e. indexicals) exhibit the same layer-crossing behavior?

Let’s take a closer look at indexicals and their possible layer-shiftiness. Let me say first of all that I am not aware of any research in this specific area and I cannot give conclusive evidence for the impossibility of Geurt-

sian de-rigidification for indexicals. That said, I can perhaps convince you that examples parallel to the ones discussed above will not easily occur with indexicals. I’ll go through them one by one. First up, the poodle named ‘Horace’ in (118). Here, a new object is introduced descriptively and then picked up by the corresponding rigid term. Note that we cannot replicate this situation with pure indexicals, because they can retrieve their own referents from the external context (“automatically”, p. 40), the descriptive introduction doesn’t affect this. We do find similar constructions with this and that:

(123) Look, I’m pointing at something. That’s is a nice specimen.

On closer examination, the anaphoric use of these demonstratives seems to go way beyond that of proper names. In fact, it’s well-known that inten-
tional indexicals have so-called discourse deictic and/or anaphoric uses in which they can pick up just about any salient sound, word, object, clause, property or proposition available from the surrounding discourse (Levinson 1983; Diessel 1999):

(124) Ligia has a poodle. I know this sounds crazy, but she loves that dog.

For lack of space I cannot attempt a comprehensive overview, classification
or analysis of all the different uses of demonstratives, nor of dedicated discourse deictics (like the former/the latter, the previous chapter). For now, I tentatively conclude that global, cross-layer binding is impossible for pure indexicals, but ubiquitous with demonstratives.

This pattern is confirmed when we go and look for non-global bindings of indexicals in constructions like (119) and (120): pure indexicals cannot be bound descriptively, cf. (125a), but other demonstratives can behave as anaphoric pronouns, (125b).

(125) a. Every time someone starts to speak, I forget what to say
    b. Everybody who owns a donkey, likes that donkey the most

Local binding is no problem for such demonstratives either (Maybe one of the donkeys was his favorite and that one escaped). Nor is accommodation, if we heap enough content in the N-part of a complex demonstrative (Did you see that Miyake movie with Sho Aikawa last night?). I’m not so sure about local accommodation, but the main observation should be that pure indexicals can’t do any of these things.46 I think we have by now firmly established that pure indexicals, proper names, demonstratives and 3rd person pronouns all display distinct resolution behavior. It seems that demonstratives, like 3rd person pronouns, are free to bind or accommodate in fr or k. Their content subsists in the opposition proximal (this) versus distal (that). The exact semantics of these features requires further research, but for now, I propose:47

(126) a. **Prel**: this \[\begin{array}{c} x \\ prox(x) \end{array}\]

---

46 With the possible exception of global accommodation of indexicals?

(i) a. I am 33 years old with blond hair and hazel eyes. I enjoy music and going to the movies.
    b. Call me Ishmael

Imagine that these sentences start a new discourse/text. The context does not (yet) provide any familiar individual that could serve as referent of the first person pronouns. On the other hand, written text brings a load of complications to any account of direct reference—think of the spatiotemporal separation between production and interpretation of the indexical token, among other things.

47 To capture textual, metalinguistic deixis (as in this chapter) we’d have to add a third, formal layer: xfr,k,form. . . . And perhaps yet another layer for cases like He went like this [strange facial expression/sounds]?
Chapter 3. A semantics of attitude reports

b. Prel: that man $\mapsto \text{dist}(x)$, man$(x)$

Proper names appear to be more tied in with the $k$-layer, lexically, but they have been known to shift occasionally. One option would be to simply generate proper names as anaphors, lexically underspecified for $k$ or $fr$, but because of the exceptional nature of descriptive proper names, I think this is not warranted. Instead I propose to analyze these cases as non-fatal violations of the layer-faithfulness constraint. But then, how about pure indexicals? Why are they always bound in the global $k$-layer.

Or are they? The term ‘shifting’ as applied to indexicals should ring a bell: monsters, 2.4.3. And indeed there is a connection between these two notions of shiftiness: monstrous shifting of indexical reference has been analyzed as involving a de-rigidification of indexicals (in 2.4.3.1 and 2.4.3.5), and the same goes for ‘bound variable readings’ of $I$ in only constructions. It seems a promising next step to try and analyze this de-rigidification as layer-shifting from $k$ to $fr$, so let’s see where that leads us.

First take Heim’s only $I$. We can derive both the sloppy (127a) and the strict (127b) readings as follows:

(a) $\Rightarrow \sim$

(127) Only I did my homework

Recall that all resolution constraints are violable, only some violations are more severe than others. What this means is that we must ensure that the layer-faithfulness constraint is not ranked too high in the final hierarchy of constraints. A possible refinement would be to split layer-faithfulness into a family of distinct, subconstraints, ranked in such a way that switching from $k$ to $fr$ comes out especially cheap.
3.3 Layered DRT

b. $\blacktriangleright \sim$

<table>
<thead>
<tr>
<th>$w_k$</th>
<th>$\text{speaker}_k(w)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{fr}$</td>
<td>$\text{ONLY}_{fr}$</td>
</tr>
<tr>
<td>$x =_{fr} w$</td>
<td>$\text{did_homework_of}_{fr}(x, w)$</td>
</tr>
</tbody>
</table>

No problem with the strict reading, as was to be expected, but for the sloppy one we have bound the second indexical presupposition (*my*) to the *fr*-labeled, quantified variable $x$. So, we’re binding a pure indexical into *fr*. There is however a big difference with the Bambi case and that is the lack of purely descriptive content associated with this $x_{fr}$; all we know is that $x = w$ where $w$ is the actual speaker in the global $k$-layer. Since equality between discourse referents is a purely extensional affair ($[x = w] \leftrightarrow \Box[x = w]$), the content of the quantified variable $x$ is fully determined by the rigid reference of $w$ in the actual context, which licenses a match between $x$ and the indexical first person presupposition $z$. The semantics of *only* as a duplex condition thus automatically handles this type of (pseudo-)layer-crossing.

The second case of apparent de-rigidification of pure indexicals was monstrous indexical shifting. Consider again our simplified shifted indexical prototype:

(128) John says that I am a hero
     ‘John says that he’s a hero’ [Pseudo-Amharic, = (50), p. 111]

It’s obvious that the Amharic *I* does not refer to the speaker of the actual context. This conspires with the previous result to make it absolutely clear that we have to give up the assumption that pure indexicals always bind in the global $k$-layer. What all analyses of monstrous indexical shifting from 2.4.3 have in common, is that they propose some kind of local binding, to a representation of the center of the reported speech act or thought. In the next section I provide a way to represent this embedded center of a belief in a $k$-layer, so that layer-shifting is not required for local binding of a first person. This move does require a special semantics of the $\text{bel}$ operator, because with the kind of operators we have seen so far, the LDRT semantics ignores $k$-layers embedded within *fr* (as shown above, p. 265). The proposed extension of $\text{bel}$ and the admittance of belief-embedded $k$ in effect amounts to a shift from intensional report semantics to monsters, from worlds to contexts, and from propositions to self-ascribed properties, just as we’ve seen before, in chapter 2. We’ll eventually return to this very example in 3.4.4.

Before we get on with belief reports, let’s take stock. In LDRT, presuppositions are labeled just like everything else. Apart from that, they are
Chapter 3. A semantics of attitude reports

represented and resolved according to the standard PA augmented with a
layer-faithfulness constraint \textit{Layer}. We distinguish two (or three) kinds of
presuppositions: descriptive presuppositions are triggered, bound and ac-
commodated in \textit{fr}, proper names and pure indexicals in \textit{k}. In between, there
are demonstratives and third person pronouns, which are not tied to either \textit{fr}
or \textit{k}. Geurts’s (1997) observations are analyzed as cases of proper names vi-
olating \textit{Layer}, i.e. crossing over to \textit{fr} and behaving like descriptive presuppo-
sitions. Pure indexicals differ from proper names in that they do not exhibit
the same kind of layer-shiftiness (which will ultimately need to be stipulated
somewhere). But there do appear to be cases of ‘bound’ pure indexicals: \textit{only}
(and related focus phenomena) and monstrous indexical shifting. Both
of these however can be tackled without allowing indexicals to switch layers,
but they do show that indexicals can be bound non-globally.

3.3.4 Belief in LDRT

3.3.4.1 Syntax

As mentioned at the end of the previous section, we need a way to put some
more structure in our belief representations and semantics. The idea is that
an LDRS representation of an ascribed belief is more or less isomorphic to
the global LDRS, i.e. with a \textit{k}-layer containing at least a center and then, 
possibly, some other discourse referents and conditions in \textit{k} or \textit{fr}. To bring
out the uniformity I’ll use the predicate \textit{center} as an abstraction of the first
person to cover the speaker or the thinker of a context depending on whether
it’s a context of utterance or a context of thought.

Take for instance a simple global context, representing the fact that the
context’s center is sick:

\begin{center}
\begin{tabular}{c}
\hline
\textit{x} \\
\hline
\textit{center}_k(\textit{x}) \\
\textit{sick}_f(\textit{x}) \\
\hline
\end{tabular}
\end{center}

(129) \hspace{1cm} [\approx \text{I am sick}]

Then, the fact that someone has a thought of the form \textit{I am sick} (as rep-
resented in (129)) is represented simply by embedding the LDRS of (129)
under a belief operator:
In other words, the global LDRS represents the context/world as seen through the collective eyes of the current discourse participants, while a BEL embedded LDRS represents how it is perceived by a particular subject.

As remarked before, this kind of representation requires a hyperintensional semantics of BEL conditions, with contexts instead of possible worlds, which is next on the agenda. Note how we are not dealing with report semantics here; though we can paraphrase contexts with sentences, we will here remain silent as to how to get from a sentence to its DRS representation. Not until 3.4 will we return to natural language attitude reports and discuss how to construct suitable PrelDRRs for simple de re/de se reports, and an accompanying extension of Res. The goal still is to give unified compositional PrelDRRs for de re/de se reports based on a relational attitude analysis, in order to ultimately derive proper, context-dependent truth conditions for all the data discussed in chapter 2.

3.3.4.2 Semantics

Since 3.3.2.4 we have a two-dimensional semantics to interpret LDRSs with k and fr, essentially providing us with Kaplanian characters. Unless stated otherwise, I’ll be using such terms as semantic value or proposition expressed as referring to the semantic evaluation of the fr-layer against rigidified k-background in context c, i.e. 3.3.2.4’s \([\varphi]^c\). As noted earlier, it would be easy to abstract away from this particular choice of layers (fr vs k) and have general L-layered content with K-layered background in a context: \([\varphi]^{K,c}_L\).

Given our 2D semantics for LDRT, we could follow Kaplan’s original analysis of belief as a relation between an individual and a full-blown character, but, as argued at length in 1.2.4.3, we might as well restrict attention to the diagonal. Remember, the diagonal was the set of contexts in which the formula, or LDRS in this case, would be true, i.e. true when evaluating the proposition expressed in context c, in the world of that c:

\[
(131) \quad \{c \in C \mid w_c \in [\varphi]^{f,c}_L\}
\]

One of the ways to think of diagonalization is as the inverse of the rigidifier
Chapter 3. A semantics of attitude reports

dthat, i.e. as an operator that de-rigidifies all indexicals and proper names in its scope. In LDRT that means a de-rigidification of the \( k \)-layer.\(^{49}\)

Next, a person’s belief set becomes a set of contexts, instead of the classic set of worlds, or, in the other extreme, the Kaplanian set of characters:

\[
(132) \quad \text{Bel} : D \times W \rightarrow \wp(C)
\]

These are the contexts that the belief subject considers actual, so, the agent(s) of these contexts correspond to the subject’s self-image, i.e. the person she believes to be.

The semantics of \text{Bel} is now entirely straightforward: believing \( \varphi \) is to think of yourself as living in one of the contexts specified by the context-property \( \llbracket \varphi \rrbracket {}_{f}^L \):

\[
(133) \quad \text{If } \xi \in \text{Dom}(f):
\]

\[
\mathcal{M} \models^{f}_{w,L} \text{Bel}_{L}^{f} \varphi \iff l \not\in L \text{ or } \llbracket \varphi \rrbracket {}_{f} \supseteq \text{Bel}(f(x), w)
\]

And that’s all. Note that we have merely followed the very standard belief semantics of self-ascribed diagonals discussed in 1.2.4.3. Later on in the next section we’ll add acquaintance relations and an extended resolution mechanism on top of this basic semantics of belief to get a handle on de re/de se reporting.

To illustrate the theory so far, consider the interpretation of our sample LDRSs (129)-(130):

\[
(134) \quad w \in \llbracket (130) \rrbracket^c
\]

iff there is a unique, smallest \( f \supseteq u_{k}(\varphi) \) \( \emptyset \) with \( \mathcal{M} \models^{f}_{c,k} \text{Colby}_{k}(w) \) and, with that \( f \), \( \mathcal{M} \models^{f}_{w,fr} (130) \)

iff there is a unique \( f \) with \( \text{Dom}(f) = \{ w \} \) and \( \mathcal{M} \models^{f}_{c,k} \text{Colby}_{k}(w) \) and, with that \( f \), \( \mathcal{M} \models^{f}_{w,fr} \text{Bel}_{fr}(129) \)

iff there is a unique \( f \) with \( \text{Dom}(f) = \{ w \} \) and \( \mathcal{M} \models^{f}_{c,k} \text{Colby}_{k}(w) \) and, with that \( f \), \( \mathcal{M} \models^{f}_{w,fr} \text{Bel}_{fr}(129) \)

iff with \( f = \llbracket w \rightarrow \text{the Colby of context } c \rrbracket : \llbracket (129) \rrbracket^f \supseteq \)

\(^{49}\)It may seem as if that could just as well have been achieved by exploiting the layers of our layered framework and evaluating that \( k \)-layer on a par with \( fr \), as in \( \llbracket \varphi \rrbracket^f_{fr,k} \). There are some differences, most notably \( \llbracket \varphi \rrbracket^f_{fr,k} \) is a set of possible worlds, while \( \llbracket \varphi \rrbracket^f_{fr} \) is a set of contexts. Additionally evaluating \( k \) and \( fr \) on a par we would lose the uniqueness otherwise built into the contextual semantics of the \( k \)-layer (with the ‘!’ operator of (106), p. 254). More precisely, if \( \varphi \) is an LDRS with a \text{center}_k condition, \( \llbracket \varphi \rrbracket^f_{fr,k} \) may contain worlds with any number of speakers and thinkers, while every \( c \) in \( \llbracket \varphi \rrbracket^f_{fr} \) has a unique center.
3.4 Belief reports in LDRT

At the end of the previous section we developed a way to represent the Lewisian belief-as-context-property-ascription in LDRT. These representations will serve as the basis for our semantic analysis of belief reports. I focus on de re/de se reports, because I have basically said all I have to say about de dicto and the de dicto/de re ambiguity. Nonetheless, let’s take stock here, picking up the report semantics where we left it in 3.2.3. In the meantime we did fix the semantics of referential terms as k-layered presuppositions, so let’s add that.

3.4.1 Recap: de dicto, de re, and de se reports in LDRT

For unambiguously de dicto reports, e.g. reports without referential/presuppositional terms in the embedded clause, I propose the following straightforward representation:

(135) Ellsworth Kimmel believes pretty girls always win  $\equiv (61a)$, p. 220

For uniformity with the de re/de se representations below, a compositional $\text{Prel}$ might add an embedded center ($\langle u_k, \text{center}(u) \rangle$) to the belief representation. Semantically, that would make no difference, because all of a person’s
belief contexts actually have a unique and salient belief center, only it’s irrelevant in the semantic interpretation of a truly de dicto belief report. Of course, this fact corresponds exactly to Lewis’s (1979a) theoretical reduction of de dicto belief to what he calls ‘de se’ (property self-ascription) by reformulating a believed proposition \( p \) as the property of living in a world where \( p \) holds, cf. 1.2.3, especially (44), p. 46.

The next example in the Ellsworth story was ambiguous between a de dicto and a de re reading. In the scopal account this came out as a pragmatic ambiguity, but after adding vivid names and acquaintance relations, we had to give that up. At the end of 3.2.3.2 I tentatively suggested some ways to re-unify the ensuing LF ambiguity, but these remained mere suggestions due to lack of space, so here I’ll just accept the ambiguity and move on. So, two preliminary structures, two readings; (136a), de dicto, following (135), and (136b) de re, following the relational attitude analysis:

(136) Ellsworth believes the prettiest girl will win \( \Rightarrow \) (61b), p. 220

\[ \begin{array}{c}
\text{Ellsworth believes the prettiest girl will win} \\
\text{BEL}_{xfr} \quad \text{prettiest}_{fr}(y) \\
\end{array} \]

a. \( \Rightarrow \) \[ \begin{array}{c}
x_{k} \\
\text{ellsworth}_{k}(x) \\
y_{fr} \\
\text{BEL}_{xfr} \\
\end{array} \]

\[ \begin{array}{c}
\text{BEL}_{xfr} \\
\text{prettiest}_{fr}(y) \\
\text{win}_{fr}(y) \\
\end{array} \]

b. \( \Rightarrow \) \[ \begin{array}{c}
x_{k} \\
\text{ellsworth}_{k}(x) \\
\text{BEL}_{xfr} \\
\text{center}_{k}(u) \\
\text{BEL}_{xfr} \\
\end{array} \]

\[ \begin{array}{c}
\text{BEL}_{xfr} \\
\text{center}_{k}(u) \\
\text{R}_{fr}(u, v) \\
\text{win}_{fr}(v) \\
\end{array} \]

\[ \begin{array}{c}
\text{BEL}_{xfr} \\
\text{center}_{k}(u) \\
\end{array} \]

---

50 I think it’s safe to assume no individual can be so deluded as to imagine she is two conscious centers at the exact same time, or none. We rely on this uniqueness of centers in 3.4.4.2.
3.4 Belief reports in LDRT

In (136b), the *de re* representation, I opted for the relational attitude analysis of 2.3.1 (applied to reports in 1.2.4.1 and adapted for DRT in 3.2.3.3), because the scopal account is deficient with respect to Ortcutt (cf. 3.2.3.2) and the original propositional vivid name fix was insufficient for capturing purely indexical distinctions as in the beliefs of Lingens or Heimson (cf. 1.2.3). To paraphrase the truth conditions of the proposed output context, there is an acquaintance relation \( R \) between Ellsworth \((x)\) and someone who’s the prettiest \((y)\), and in all contexts compatible with Ellsworth’s beliefs the center \((u)\) is R-acquainted with someone \((v)\) who wins. As discussed before, there’s a number of restrictions on \( R \), such as uniqueness of the second argument and vividness, but we’ll leave those out for readability. Note by the way that in this particular resolution of the *de re* PrelDRS I chose to accommodate the presupposition triggered by *the prettiest*, but if it was already salient common ground that Ellsworth is the prettiest, we could have bound it to \( x \) instead. On the other hand, if that were the case it would have been much more efficient to just use a pronoun (138) and avoid the whole *de dicto/de re* ambiguity. In any case, semantically speaking it doesn’t make any difference whether we bind or accommodate, because in the actual, external contexts of utterance that come with our scenarios Ellsworth = Arata = the prettiest.

The following *de re* report brings nothing new:

\[
(137) \quad \text{Ellsworth believes of Arata Suggs that she’ll win} \quad \text{[= (61c), p. 220]}
\]
Again, if in the discourse context it was a known fact that Arata is Ellsworth we have the option of binding, but then, why would I even use (137)? Why confuse my audience by using two different proper names, where a simple anaphoric pronoun would have sufficed? Surely that would fly in the face of the conversational maxims, the hierarchy of referential/anaphoric expressions of (5), p. 357 in particular. Another possibility, if this is the audience’s first encounter with the name Arata, we might well consider a layer-switch to fr, i.e. a kind of reinterpretation of this Arata as someone called Arata (cf. 3.3.3.2). So, depending on the broader conversational background there is the choice to put y in either k or fr.

Now, the pronoun case, where things should get interesting:

(138) Ellsworth believes that she’s going to win

The big payoff of the relational attitude analysis: no de re/de se ambiguity in co-referential pronoun reports. Note first that the representation above is compatible with a mistaken identity de re belief of Ellsworth’s as in the TV
scene B from (59), p. 219. To see that, take \( R \) to be the relation of ‘seeing some Arata on TV’ then \( R \) holds between Ellsworth and herself (in \( c_B \)), and from her defeatist remark in \( c_B \) we can deduce that she believes the person she is acquainted with through that particular \( R \) will win. So, (138) is true in \( c_B \). Taking \( R \) to be equality, we see it’s equally true in \( c_C \), where the belief is purely \( de \ se \), in accordance with the originally Lewisian claim that pure \( de \ se \) is a subtype of \( de \ re \), viz. \( de \ re \) under the acquaintance relation of equality.

Which brings us to our final Ellsworth-sentence, the unambiguously pure \( de \ se \) belief report:

\[
(139) \quad \text{Ellsworth believes PRO to be on the winning side} \quad \begin{array}{|c|}
\hline
x_k \\
\hline
\text{ellsworth}_k(x) \\
\hline
\hline
u_k \\
\text{center}_k(u) \\
\text{win}_k(u) \\
\hline
\end{array}
\]

[\text{cf. (76), p. 233}]

Its truth conditions require that Ellsworth has a belief corresponding to the set of contexts whose center will win, i.e. the pure \( de \ se \) belief she expresses as “I will win” in \( c_C \). There is an important connection between this LDRS and the more complex relational one in (138) and that is that we can get from (138) to (139) by substituting \( = \) for \( R \), basically just applying the reasoning that shows pure \( de \ se \) is relational \( de \ re \) belief under equality.

### 3.4.1.1 Known issues

I will not touch \( de \ dicto \) anymore, nor the \( de \ dicto/de \ re \) interface, which ideally would deserve further attention, as pointed out above. I will only consider the overlapping \( de \ re \) and \( de \ se \) modes. The proposed analysis basically implements Cresswell & von Stechow’s (1982) relational analysis in LDRT. It works for simple \( de \ re \) belief reports and Orcutt cases, and it gives us a nice reduction of pure \( de \ se \) as a special case of relational belief. On the other hand, we also inherit the bugs that came to light in chapter 2. I here list these issues briefly and then spend the rest of the thesis extending the basic relational formalism described above.

The first issue discussed was context dependence. In 2.3.2 it was suggested that existential quantification over acquaintance relations is inferior to an account that selects a specific acquaintance relation from the context or at least contextually restricts the acquaintance quantification. I criticized the proposed contextual alternatives for their lack of a formal theory about this
pragmatic restriction. The dynamic framework of DRT seems quite suited to this task, so the first thing I’ll do, in 3.4.2, is to formulate an extension to the resolution algorithm to handle the contextual resolution of second-order variables like our $R$. I’ll illustrate how we get the benefits pointed out by Aloni and Abusch but now properly integrated in a pragmasemantic framework that takes context dependence as its starting point. Later, in 2.4.2, I showed that certain quantified belief reports, with *every* and *only*, cause problems for a strictly wide-scope account of contextualized acquaintance. I will propose a remarkably simple extension to the LDRT variant of 3.4.2, henceforth known as Acquaintance Resolution, in order to account for these data.

Second, in 2.3.3 we discussed how compositional the relational analysis of *de re/de se* is. We saw that it minimally requires separating the *res* from the ascribed predicate, thus preventing a fully Montagovian analysis of belief embeddings. It was also remarked that contextualizing the acquaintance relation makes things worse, compositionality-wise (though at first sight they may seem slightly better). Transposed to LDRT, the first objection still holds: we need to know which embedded elements are to be read *de re*, because only for those will we introduce an acquaintance variable. However, the second observation, about contextuality, is no longer valid, since compositionality in this framework only applies to the construction of PrelDRSs, not outputs. In (L)DRT+PA, no output is derived fully compositionally anyway, but the use of underspecification in PrelDRSs makes it relatively easy to formulate a highly compositional construction algorithm (Prel). It’ll be easy to see how this is achieved in our basic analysis of 3.4.2 as well as in the later extensions.

Finally, as discussed at length in 2.4 there is a strong anti-reductionist movement advocating independent *de re* and pure *de se* LFs. It started with Chierchia’s (1989) observation that there are ways to report just pure *de se* beliefs, for instance with infinitival complements, as in (139). This was taken as evidence that *de re/de se* is not a homogeneous class and that we should distinguish separate LFs for *de re/impure de se* on the hand and pure *de se* on the other. Chierchia himself even went so far as to claim that co-referential pronoun reports like (138) are syntactically ambiguous in this respect. There are some arguments involving elided reports, but these were shown to be inconclusive at best, cf. 2.4.1.4. In 2.4.2.2 we considered a stronger argument in favor of the Ambiguity Thesis. The main observation was that embedding (138) under *only* allows a reading that doesn’t correspond to either narrow or

\[51\] Ultimately, the goal should be to put the creation of the acquaintance relation in the resolution algorithm, so that we get a situation similar to the scopal account where narrow resolution gives *de dicto*, wide scope *de re*. Such a solution would *ipso facto* solve our problem with the *de dicto/de re* ambiguity too.
3.4 Belief reports in LDRT

wide scope quantification/resolution of an acquaintance relation, but which
is easily explained if we assume that there is a distinct, dedicated pure de se
LF. This poses one of the tougher challenges for our de re/de se unificationist
attempt, but it will be met, in 3.4.3.

An alternative, which I refer to as the hybrid account of de re/de se,
consists in assigning a dedicated pure de se LF to infinitival reports, while
keeping a general de re/de se relational LF for the pronoun reports. In fact,
this is more or less what we did in (139) above. As pointed out there, we
can still maintain a somewhat weakened version of the reductionist idea that
pure de se is just de re under the particular acquaintance of equality, if only
semantically. On the downside, since the acquaintance relation is precisely
not underspecified for infinitival reports, as it is for co-referential pronouns, it
seems we have to give up on the idea of a unified preliminary representation
for the two syntactic structures. This in turn means we can give up PRO,
which was recruited to unify the syntactic structures of the embedded clauses
(into the subject-predicate mold). Since I’m not a generative syntactician,
I can live with the loss of this invisible pronoun. All in all, it seems a fair
proposal.52

But then there’s other data, reports that are similarly interpreted purely
de se, but where the complement does contain an overt pronominal subject.
First, there are the logophors discussed in 2.4.1.2. These are a class of third
person pronouns whose effect is not simply co-reference with the matrix sub-
ject, but an unambiguously pure de se attitude ascription to that subject.
Now it seems there is a definite advantage to having an account of PRO’s
pure de se semantics, because if we did, we could use that as a semantics
of PRO and logophors uniformly. Finally, there’s the phenomenon of shifted
indexicality of which the Amharic first person has come to be the prime ex-
ample. It’s commonly assumed that the shifted reading of such an I, where
it refers not to the actual center, but to the center of the reported belief,
corresponds to a pure de se ascription. Following Schlenker and others I pro-
pose the obvious, viz. that in such cases the Amharic I, a k-presupposition,
is bound locally to the embedded center. The wide scope resolution should
correspond to the other, English-type, reading that is de re about the actual
speaker. The difficulty is to get these two distinct readings to come out just
by resolution from a single de re/de se PreIDRS. In 3.4.4 I have planned a
considerable overhaul of the system to achieve this feat. But once we have
it, we also have a way to represent PRO and logophors: just assign them the

52To be sure, we’d have to check some of the more subtle data involving ellip-
sis/anaphorization of report complements (2.4.1.4), but since I currently lack a sufficiently
precise theory of abstract anaphora in LDRT, we’ll leave that for another occasion.
same, first person $k$-presupposition as Amharic $I$. Unfortunately this would overgenerate wide scope readings, so in the end we are still left with something like Schlenker’s stipulative typology of indexicals to make sure that PRO and logophors are always bound locally.

This concludes our schedule for the remainder of the thesis. If everything goes according to plan, I will have provided the first de re/de se report semantics that takes the context dependence of acquaintance relations seriously and in doing so unifies the de re and pure de se modes of reporting belief. More specifically, this semantics, formalized in LDRT, will be able to take on any kind of de re or de se report of the form $NP$ believes that $NP$ VP, including Ortcutt cases, co-referential pronoun reports and quantified versions thereof—in mistaken-identity and mixed scenarios—, infinitival pure de se reports, logophors, and shifted indexicals.

### 3.4.2 Acquaintance Resolution

In the previous section we saw logical forms with existentially quantified acquaintance relations, i.e. $x$ believes something de re of $y$ iff there exists a vivid acquaintance relation between $x$ and $y$ such that, etc.. As we have argued elsewhere, it might be a good idea to restrict the search for acquaintance relations to the context: $x$ believes de re of $y$ iff the context provides a salient acquaintance relation that holds between $x$ and $y$, etc. There are a number of ways to make this formally precise; the one I propose involves the resolution of a second order variable, $R$, by unification of the $R(x, y)$ condition with the context represented in the input LDRT.

First, I’ll talk about unification and how we can use it to model the context dependence of acquaintance, then we’ll illustrate Acquaintance Resolution with the Ortcutt example and the coreferential pronoun report about Ellsworth.

#### 3.4.2.1 Higher-order unification

So, what is unification exactly? In general, unification, applied to a set of expressions, is the process of finding a unifying substitution for those expressions, i.e. a substitution of terms for free variables that makes the expressions equal. Take for example the first-order predicate logical expressions $\text{walk}(x)$ and $\text{walk}(j)$. We’ll consider only unifications of two expressions, in which case it makes sense to talk about unifying their equation. Obviously, though, the equation in question cannot be the relation of equality commonly denoted by the 2-place relation symbol $=$ in first-order logic. Rather, it denotes a kind of formal, syntactic equality between terms themselves. We’ll return
3.4 Belief reports in LDRT

to the exact interpretation of this ‘formal equality’ shortly, for now, we’ll just use a different symbol, \( \doteq \). Returning to the first-order example: (140) shows how we’ll represent this unification problem, along with two possible solutions:

\[
(140) \quad \text{walk}(x) \doteq \text{walk}(j)
\]

a. \[
\begin{array}{c}
  \text{walk}(x) \\
  x \mapsto j
\end{array}
\]

b. \[
\begin{array}{c}
  \text{walk}(x) \\
  x \mapsto j \\
  y \mapsto z
\end{array}
\]

The first, \((140a)\), is indeed a unifier because applying it to (both sides of) the unification problem yields a trivial truth (since \( \text{walk}(x) [x \mapsto j] \doteq \text{walk}(j) \)). The same holds for \((140b)\), but this one is not minimal, so to speak, because it contains the first unifier in the sense that in can be decomposed as the composition of the unifier \((140a)\) and \([y \mapsto z]\). We say that \((140a)\) is the most general unifier of the equation. It is a well-known fact that first-order unification problems that can be unified always have a unique most general unifier. Moreover, there’s an algorithm for deciding whether unification is possible and, if so, retrieving the most general unifier. In the following we’ll consider only most general unifiers, and we often just use the terms ‘unifier’ or ‘solution’ for them.

Now, in a higher-order language, such as the typed lambda calculus of 2.B, we can formulate the same kind of unification problems, with higher-order variables:

\[
(141) \quad X(j) \doteq \text{walk}(j)
\]

\[
\begin{array}{c}
  X \mapsto \text{walk}
\end{array}
\]

Assuming \( X \) and \( \text{walk} \) are of type \( \text{et} \) (1-place predicates), this is the obvious resolution. For details regarding higher-order unification I refer the reader to Huet (1975).

But now consider again the question of what \( \doteq \) really means. As we said, it’s not the purely semantic notion that requires the extensions/intensions to be equal, for we don’t want to count \( \text{walk}(j) \) as equal to \( \text{walk}(j) \land [\text{rain} \lor \neg \text{rain}] \). On the other hand, what we do count as equal are \( \text{walk}, \lambda x.\text{walk}(x) \), and \( \lambda y.\text{walk}(y) \), for instance. In other words, \( \doteq \) stands for the equivalence relation of \( \alpha\beta\gamma\)-interconvertability of expressions, as defined in 2.B.5. In words, two expressions of the same type are \( \doteq \)-equal if we can get from one to the other by renaming bound variables, ‘\( \lambda \)-style’ function application, and ‘dummy-abstraction’ (as exemplified by \( \text{walk} \leadsto \lambda x.\text{walk}(x) \)). Applied to \((141)\), this means there is another solution: \( [X \mapsto \lambda x.\text{walk}(j)] \). The
Lambda term denotes the pseudo-property that holds of someone iff John walks, so when we substitute it for \( X \) and apply it (vacuously) to \( j \), the result is \( \text{walk}(j) \). This substitution is not equivalent to the one in (141), which may now also written as \( [X \mapsto \lambda x. \text{walk}(x)] \), but it is no less general.

So, we’ve lost the uniqueness result for higher-order unification. How about decidability? In general, we lose that too, but for specific subclasses, decidability has been proven. Fortunately, we will be working in such a well-behaved subclass, called second-order matching. In particular, we’re interested in the resolution of second-order relation variables, acquaintance relations, for which it’s actually very easy to enumerate all solutions.

### 3.4.2.2 The resolution of acquaintance

After this theoretical interlude we return to the problem of the contextual resolution of acquaintance relations. My idea is to take the statement that \( R \) holds between the subject and the \( \text{res} \) of the report and then search the context for appropriate conditions relating these entities to form a second-order unification problem. By unifying the formal equation we resolve the variable \( R \). Because this is higher-order, there’ll be more than one solution, but the number will be finite and we can formulate a constraint to narrow it down to one. Applying the optimal unifier to the entire LDRS under construction gives the desired output.

An example will make this clear. Consider the following \( \text{de re} \) report, uttered in a discourse context where Colby and Ligia are known to be neighbors:

(142) Colby thinks Ligia is cool

\[
\begin{array}{c|c}
\hline
x_k & y_k \\
\hline
\text{colby}_k(x) & \text{ligia}_k(y) \\
\text{neighbor}_k(x, y) & \\
\hline
\end{array}
\]

\[
\begin{array}{c|c}
\hline
u_k & v_k \\
\hline
\text{center}_k(u) & \text{R}_k(u, v) \\
\text{cool}_k(v) & \\
\hline
\end{array}
\]
At this point we want to formulate an appropriate unification problem, i.e. we need to isolate a salient relation between $x$ and $y$. Moreover, this relation has to describe some kind of vivid link between $x$ and $y$, because it will have to serve as an acquaintance relation (cf. 1.2.2.4 for some discussion about what constitutes a valid relation of acquaintance). It seems the \textit{neighbor} relation is the best candidate here (other options are considered below). Note that, because simply being neighbors does not necessarily amount to any vivid perceptual relations, we shall assume that in this case the condition $\text{neighbor}_f(x, y)$ is really an abbreviation of the various interactions that constitute the actual relationship between them. Next, let’s leave the layering out of the equation, we’ll just unify bare DRS conditions and then later, when applying the substitution, we’ll put everything at the designated layer. With these preliminary simplifications in place, we get:

\[
\begin{align*}
R(x, y) & \vdash \text{neighbor}_f(x, y) \\
\text{a. } & R \rightarrow \lambda s \lambda t[\text{neighbor}_f(s, t)] \\
\text{b. } & R \rightarrow \lambda s \lambda t[\text{neighbor}_f(s, y)] \\
\text{c. } & R \rightarrow \lambda s \lambda t[\text{neighbor}_f(x, t)] \\
\text{d. } & R \rightarrow \lambda s \lambda t[\text{neighbor}_f(x, y)]
\end{align*}
\]

It’s easy to see that these are the only solutions. Now we have to pick one and apply it to the LDRS under construction. Since both the first and the second argument of $R$ vary across the LDRS ($R_f(x, y) - R_f(u, v)$) it makes sense to pick a solution that takes both arguments into account, i.e. (143a). If we took (143b) for instance, the second argument, the $t$, would just be swallowed up and replaced with a $y$. In (143d), both arguments are dummies, in the sense that it would verify $R(x, y) \vdash R(u, v)$, which would ultimately yield a totally undesirable output. I thus propose a constraint on second-order acquaintance resolutions to the effect that both arguments to the acquaintance relation are relevant in the sense just described.
So, (143a) it is. Continuing (142):

\[
\begin{array}{c}
\text{colby}_k(x) \\
\text{ligia}_k(y) \\
\text{neighbor}_f(x,y) \\
\hline
\text{R}(x,y) \\
\hline
\text{center}_k(u) \\
\text{R}(u,v) \\
\text{cool}_f(v) \\
\end{array}
\]

The reading we predict ascribes to Colby the belief that his neighbor’s cool, which is true and seems like a plausible reading of the sentence in the given context.

Perhaps, it would be even better if we took the property of being called Ligia along with the neighbor relation, i.e. take (145) instead of (143):

\[(145) \quad \text{R}(x,y) \doteq [\text{ligia}_k(y) \text{neighbor}_f(x,y)] \]
3.4 Belief reports in LDRT

a. \([R \mapsto \lambda s \lambda t[\text{ligia}_k(t) \text{neighbor}_f(s, t)]\]

b. \([R \mapsto \lambda s \lambda t[\text{ligia}_k(t) \text{neighbor}_f(s, y)]\]

\vdots

h. \([R \mapsto \lambda s \lambda t[\text{ligia}_k(y) \text{neighbor}_f(x, y)]\]

If we take the non-trivial solution, (145a), we end up with a reading that ascribes to Colby the belief that the neighbor he calls Ligia is cool. Depending on the rest of the context I think this is probably a slightly more accurate reading for most utterances of (142). After all, it makes sense, pragmatically, to only use the name Ligia in a report to reflect part of the description that played a role in the reported thought.

But what about just taking the proper name condition?

(146) \(R(x, y) = \text{ligia}_k(y)\)

a. \([R \mapsto \lambda s \lambda t[\text{ligia}_k(t)]\]

b. \([R \mapsto \lambda s \lambda t[\text{ligia}_k(y)]\]

The least trivial solution would be (146a), but does this still count as an acquaintance relation? First, it doesn’t involve perception, and second, it’s not even really a relation. On these grounds we can safely discard it as a possible resolution of acquaintance. If we were to accept it, just for the sake of argument, we’d get an output that says Colby believes the person called Ligia is cool, i.e. a purely \textit{de dicto} reading of the proper name (actually equivalent to what we’d get from a narrow scope resolution in a simple scopal account as in 3.2.3.1). This neatly echoes the idea that \textit{de re} belief requires a proper relation of acquaintance. On the other hand, this kind of derivation might eventually be exploited to unify \textit{de dicto} and \textit{de re}, though not here.

3.4.2.3 Applications: Ortcutt and mistaken identity

Let’s go through another example, 1.2.2’s version of Quine’s (1956) Ortcutt case, in which Janell is acquainted with Ortcutt under two different guises. The main aim of the derivation is to further clarify the workings of Acquaintance Resolution. At the same time it’s a demonstration of the system’s flexibility, providing a novel solution to Quine’s puzzle of \textit{de re} belief and reports.

The scenario is as described in (70)-(71), p. 227, which gives a DRS representation of an appropriate input context modeling the common ground at the point of the report, as well as a description of the actual state of affairs that the speaker is trying to report. We start with the input context and see what happens if we first interpret (147) and then the opposite, though
crucially not inconsistent, ascription (147).

(147) Janell thinks Ortcutt is dumb

\[
\begin{array}{c|c}
 x_k & y_k \\
\hline
 \text{janell}_k(x) & \text{ortcutt}_k(y) \text{namos}_f(y) \\
 \text{admire}_f(x,y) & \text{write}_\#042_f(y) \text{referee}_f(x,y)
\end{array}
\]

Once anaphora and other presuppositions are taken care of it’s time to resolve the free variable, \( R \). We are looking for contextual information that will eventually give us a real two-place relation for \( R \), so we choose a set of conditions from the context containing both an \( x \) and a \( y \). This means we consider only subsets of the conditions in the input LDRS containing either \( \text{admire}(x,y) \) or \( \text{referee}(x,y) \). To further narrow it down there’s the global resolution constraint requiring consistent and coherent outputs, i.e. pragmatics. The obvious candidate gives the following unification problem and solution:

(148) \[ R(x,y) \doteq \left[ \text{write}_\#042_f(y) \text{referee}_f(x,y) \right] \]
\[ [R \mapsto \lambda s \lambda t [\text{write}_\#042_f(t) \text{referee}_f(s,t)]] \]

I say this is the obvious unification because it yields a highly plausible output if we apply it to (147):
### 3.4 Belief Reports in LDRT

<table>
<thead>
<tr>
<th>(x_k y_k)</th>
<th>janell_k(x)</th>
<th>ortcutt_k(y) famous_f(y)</th>
<th>admire_f(x, y)</th>
<th>write_#042_f(y) referee_f(x, y)</th>
</tr>
</thead>
</table>

In other words: uttered in the context sketched above, the sentence in (147) is predicted to mean that Janell had a thought of the form “the person who wrote \#042, which I refereed, is dumb.”

Now the conversation continues and at a certain point, most likely after it has been brought up (again) that Janell is such a great admirer of Ortcutt’s work, someone utters (150). We can continue the derivation as follows:

\[
\begin{align*}
\text{(150) & Janell thinks Ortcutt is brilliant} \\
\text{[=}(70b), \text{p. 227]} \\
\end{align*}
\]

<table>
<thead>
<tr>
<th>(x_k y_k)</th>
<th>janell_k(x)</th>
<th>ortcutt_k(y) famous_f(y)</th>
<th>admire_f(x, y)</th>
<th>write_#042_f(y) referee_f(x, y)</th>
</tr>
</thead>
</table>

\[
\begin{align*}
\text{BEL}_f & \Rightarrow \ldots \\
\end{align*}
\]

\[
\begin{align*}
\text{R}(x, y) & \Rightarrow [\text{ortcutt}_k(y) \text{ famous}_f(y) \text{ admire}(x_f, y)] \\
& \Rightarrow [R \rightarrow \lambda s \lambda t [\text{ortcutt}_k(t) \text{ famous}_f(t) \text{ admire}_f(s, t)]]
\end{align*}
\]
Chapter 3. A semantics of attitude reports

We see how the second sentence’s acquaintance relation picks up the other $x, y$-relation, the one that connects Janell to Ortcutt as the famous professor she admires. This choice can be seen as driven by the consistency constraint, for if we were to take the same relation as before we’d end up with contradictory belief sets. By contrast, the reading derived above has Janell self-ascribe a belief set in which the center admires someone named Ortcutt who’s famous and brilliant.

In conclusion, by providing enough context and choosing the right conditions resolving the $\text{Prel}$-generated acquaintance variables, we can derive the proper non-contradictory readings for the reports of the Ortcutt example: Janell believes the person who wrote abstract #042, which she refereed, is dumb, while she also believes that the famous Ortcutt she admires is brilliant. In addition, the input and output LDRSs contain the information that Ortcutt is both the author of #042 and the famous professor admired by Janell.

We now return to the Ellsworth test cases. It should be pretty clear how the derivations run under Acquaintance Resolution, so let’s skip to the interesting case, (61d), the coreferential pronoun report in mistaken identity and the pure $de se$ contexts, B and C. We start with an utterance of (151) in a mostly empty discourse context, containing just a salient Ellsworth, i.e. the common ground component we’ve been using for dynamic interpretation with respect to A and C:

(151) Ellsworth believes she’s going to win

\[\text{\(=\text{(61d)}\), p. 220}\]
3.4 Belief reports in LDRT

And then we need to find a relation between \( x \) and \( x \) in the context. But there’s nothing in the context, except the unary predicate of being called Ellsworth, which, we showed earlier will not do if we are to derive de re readings. I propose to equate \( R(x, x) \) with \( x = x \), or rather, to unify \( R \) with \( = \), the relation of equality. This should be allowed, for even if it is not explicitly represented, \( x = x \) nonetheless holds in every context. Also, I assume, with e.g. Lewis (1979a) that equality counts as a relation of acquaintance, capturing the idea that every individual is primarily acquainted with itself as itself.

In any case, in the relational analysis that inspired our LDRT semantics of de re, pure de se is de re under the acquaintance relation of equality. To be sure, let’s verify that taking equality as the relation of acquaintance will give us a first person de se belief.\(^{53}\)

---

\(^{53}\)the last step, getting rid of the \( v \) variable
Chapter 3. A semantics of attitude reports

(152) \( \sim R(x, x) \triangleq x = x \sim [R \leadsto \lambda s \lambda t [s = t]] \sim \)

\[
\begin{array}{|c|c|}
\hline
x_k & \text{ellsworth}_k(x) \\
\hline
\text{BEL}_{x_f} & \text{center}_k(u) \\
\hline
u_k & v_f \\
\hline
\end{array}
\]

Apparently, for coreferential reports (where the subject is reported to have a \textit{de re} belief about herself) we can always derive a pure \textit{de se} reading, by resolving \( R \) to equality. However, if the context provides other information that presents the reportee to herself in a more indirect way, we may go with that as well, leading to an impure \textit{de re} belief about the subject herself. This is what happens with the pronoun report (151) when uttered in B, the mistaken identity scenario. Of course, the default equality option is still there, but now the pure \textit{de se} reading would be false, Ellsworth does not think “I’ll win” in B, and therefore we search through the context a little better and find the fact that \( x \) sees \( x \) on TV:

(153) Ellsworth believes she’s going to win \([=(61d), \text{p. } 220]\)

\[
\begin{array}{|c|c|}
\hline
x_k & \text{ellsworth}_k(x) \\
\hline
\text{arata}_k(x) & \text{BEL}_{x_f} \\
\hline
\text{see on tv}_f(x, x) & \text{center}_k(u) \\
\hline
R(x, x) & u_k v_f \\
\hline
\end{array}
\]

\( \sim R(x, x) \triangleq [\text{arata}_k(x) \text{ see on tv}_f(x, x)] \sim [R \leadsto \lambda s \lambda t [\text{arata}_k(t) \text{ see on tv}_f(s, t)]] \)
So, if we opt for this resolution of $R$ we obtain a reading in which Ellsworth believes something she’d paraphrase as: “the person called Arata I see on TV will win”, which corresponds to the (true) *de re* reading that, following Kaplan and others, we managed to squeeze out of pronoun reports like (61d), p. 220 (repeated as (151) below).

The main point I want to make here is that our model can actually account for the effort it takes to convince non-philosophers of the existence of this mistaken-identity reading of reports with coreferential matrix and embedded subjects. The only thing we need to add is that equality is the default resolution of $R$ if an $R(x, x)$ is given. This means we always try equality, i.e. the pure *de se* reading, first, and only if that yields an inconsistent or otherwise pragmatically unfit output do we turn to the context. Assuming that this pragmatic check and recalculation with additional context searching somehow correlates with actual processing on the part of the hearer, we get at least the first step towards an explanation of the perceived awkwardness of an impure reading for such reports. It also follows that mistaken-identity *de re* readings only occur if the indirect acquaintance (through mirror, picture, or TV) is sufficiently salient in the discourse context of the report. This is very much what we perceive to be the case: the pronoun report is not just true if some acquaintance relation exists, as the classical relational analysis would have it, rather, the default reading is pure *de se*, which may be overruled in case of significant contextual pressure and the availability of an alternative acquaintance. I consider this one of the major advantages of my LDRT Acquaintance Resolution over the classical relational analysis (or the LDRT version demonstrated in (138)).

Compared to pure *de se* separatist accounts, like Chierchia’s, the obvious advantage of Acquaintance Resolution is that we do not require a syntactic ambiguity. We assign a single, underspecified *de re* PrelDRS to (61d), and the contextual, pragmasemantic resolution mechanism does the rest. Note
that the pure and impure readings that come out of the resolution process eventually are truth conditionally distinct, in fact there are infinitely many distinct outputs, varying with the context, of which pure *de se* is but one, though privileged, case.

So far, our account seems quite promising, a sound and elegant formalization of 2.3.2’s suggestion of contextualizing acquaintance. In addition, we retain a compositional PrelDRSs construction, at least as compositional as any acquaintance-based theory can be (i.e. we still need to specify whether the report is to be interpreted relationally, and which are the *res*). As pointed out in chapter 2 however there are a number of examples where contextualized acquaintance seems to give the wrong predictions. First, there is embedding under quantifiers, where some data seems to point in the direction of the classic quantified acquaintance framework, while other examples pose problems for both contextualized and quantified acquaintance. I extend the Acquaintance Resolution framework to capture all of these data in 3.4.3. A more substantial extension is required for another set of data adduced against *de re/de se* unification in general, those involving unambiguously pure *de se* reports and shifted indexicals. We turn to these in 3.4.4.

### 3.4.3 Quantifiers

In this section we are concerned with two types of quantificational embedding, first universal quantification, then quantification under *only*. When placed in a mixed scenario where some have a pure and others an impure *de se* attitude about themselves, these pose interesting challenges for our acquaintance-based framework. We’ll take up the minimal mixed scenario of 2.4.2:

(154) \[ c_D \text{ Ellsworth, aka Arata: “That Arata person on TV is probably going to win”; Shavonne: “I’m gonna win.”} \] \[ \approx_{D, \text{1, p. 132}} \]

\[ \varphi_D \]

\[ \begin{array}{|c|}
\hline
x_k & y_k \\
\hline
\text{ellsworth}_k(x) & \text{arata}_k(x) \\
\text{see_on_tv}_g(x, x) & \text{shavonne}_k(y) \\
\hline
\end{array} \]

We could follow the derivations of the previous section to show that Acquaintance Resolution gives the desired result for unquantified reports:

(155) \[ \text{a. Shavonne thinks she’s going to win} \] \[ \approx (80a), \text{p. 132} \]

\[ \text{b. Ellsworth thinks she’s going to win} \] \[ \approx (80b), \text{p. 132} \]
3.4 BELIEF REPORTS IN LDRT

Now let’s see what happens with quantifiers.

3.4.3.1 Universal quantification

Zimmermann’s argument in favor of a unified de re/de se semantics (à la Kaplan (1989), but in this respect there’s no difference with the classical relational analysis) is based on the following report he judges true in mixed scene D:

(156) Both women think they’re going to win

What does Acquaintance Resolution predict? Well, using a duplex condition (3.1.4.1) for the quantifier, our compositional Prel assigns (157) as the preliminary LDRS:

This is not entirely straightforward, first, how to interpret the plural morphology of women (and they)? The quantified variable z is semantically singular, ranging over individuals, not sets, so we’d expect women to be interpreted as a singular. An in-depth treatment of this issue is Kamp & Reyle’s (1993) analysis of dependent plurals.

My account is simpler. I’ll say that the plural morphology here is but a surface phenomenon without a semantic correlate. This is corroborated by the fact that other quantifiers can say the exact same thing with singulars, evidently depending on some lexical feature of the particular quantifier. Take a simpler example:

(158) a. Every student did his homework
    b. All students did their homework
The truth conditions surely are the same, it’s just that *every* is lexically specified to take a singular, while *all*, like *both* in (156), demands a plural.

Now how about the morphology of the pronoun later in (156), or those in (158)? How should we interpret and represent these? Obviously, in the most salient readings of (156) and (158) the pronoun will be bound by the quantifier, i.e. bound by the quantified variable in the restrictor’s universe. Since we just decided to ignore the plurality of the quantifier’s restrictor, we better also delete that particular feature from the pronoun or we’ll be unable to derive the intended reading.

The situation seems to fall under the general phenomenon of *feature deletion (and consequent binding) under morphological agreement*, cf. 2.4.3.4. However, I will not adopt the extended feature deletion mechanism of Heim (ms) or von Stechow (2002). Instead I intend to do as much as possible with straightforward semantic and pragmatic mechanisms, simply interpreting grammatical features as they appear on the surface. As a matter of methodological principle I try to avoid morphosyntactic movement and deletion at all costs, unless of course we have a phenomenon like this, where it’s absolutely clear that there is no semantic explanation. In fact, one of my main objectives was to show a way to interpret belief-embedded pronouns pragmasemantically, without feature deletion.

Thus, I am forced to maintain that the apparent agreement phenomena encountered in quantificational environments constitute a separate issue from those in belief reports. As independent evidence in support of this claim consider (159), which shows that the features of quantifier-bound pronouns are subject to cross-linguistic variation without semantic consequences (save that, for some, the feminine gender feature does add a domain restricting presupposition):

(159)  {Geen/ieder} van ons heeft {zijn/haar/*ons} huiswerk gedaan
       {none/each} of us has {his/her/our} homework done
   ‘{None/each} of us did {?his/?her/our} homework’

[German vs English]

Apparently, the choice of bound pronoun forms differs not only from quantifier to quantifier, but also from language to language (and dialect to dialect, if you’d classify the non-sexist (and anti-sexist) varieties of English as such)—strong evidence for its a-semantic nature. Contrast this with the very strict number, person and gender agreement between matrix subject and coreferential embedded pronoun in pure *de se* reporting, even to the point where it appears to contradict the semantics:

(160)  John, a transsexual, wants to become a woman, and then he hopes
3.4 Belief reports in LDRT

The ‘paradox’ of (160) is the masculine feature of the embedded his which, on the most natural reading (pure de se), refers to the belief center, who in this context is a post-operative female. For de se separatists like Chierchia or even Schlenker this is a problem that eventually leads to the introduction of an across the board feature deletion under agreement mechanism. Acquaintance Resolution, or any other relational analysis of de re/de se, accounts for this observed feature agreement semantically, by binding the anaphoric pronoun globally, before resolving the acquaintance relation to equality. In any case, pure de se readings of (third person) pronouns in reports co-occur with strict feature agreement between the embedded pronoun and its global ‘antecedent’, independently of a particular choice of attitude verb, language or dialect.

Another difference between quantificational and attitudinal pronominal agreement is the first person. An I in a belief report obviously allows for pure de se (‘bound’) readings in case of agreement with a matrix first person, but, as noted before, bound readings of I are not so easy to obtain with quantifiers. In fact, Schlenker (2003:90) claims that besides attitude reports, only ellipsis constructions and focus-sensitive constructions like with only seem to be able to bind I through feature deletion. Conclusion: despite some prima facie similarities, there is enough reason for assuming a number of distinct phenomena that create superficially similar binding under agreement appearances. I’d assume up to four separate mechanisms: (i) For ellipsis, I’d prefer Dalrymple et al.’s (1991) account based on higher-order unification which has no need for feature deletion in the analysis of I did my homework, and so did Janell. This unification-based approach was applied to focus by Pulman (1997), and this would probably suffice to handle the first person binding of Heim’s only I did my homework. However, (ii) as shown in 3.3.3.2, my analysis of only as a quantifier in combination with I as a k-presupposition already accounts for both the strict and the sloppy readings without feature deletion or higher-order unification. (iii) For de re reports I use Acquaintance Resolution, which incorporates a relational, wide scope analysis of embedded pronouns that automatically derives the agreement with the matrix subject on pure de se readings. What’s left is (iv) the class of pronouns bound by quantifiers for which I’m forced to rely on some a-semantic feature deletion rule for number features (possibly with a restricted extension to gender and

54 Schlenker goes on to argue against Heim (ms) and von Stechow’s (2002) extended feature deletion mechanisms by claiming that “[a]ttitude reports do not appear to form a natural class with either [ellipsis] or [focus], and thus it is not obvious that a unification can be achieved.”
person). The upshot of this rather lengthy intermezzo on quantificational feature agreement is that from now on my PrelDRSs will freely disregard number features of pronouns in quantificational environments.

Back to our quantified belief report then. I assume the syntactic construction algorithm has properly deleted the plural features by the time the context is added. Resolution proceeds by first binding the pronoun to the quantified variable $z$:

At this point we want to single out an appropriate part of the context to equate $R(z, z)$ to. In cases of self-acquaintance such as this, we agreed that
equality is the default. In fact, here it’s the only possibility, for the essential ingredient \( z \) doesn’t even occur in the global LDRS. Equality it is then:

\[
\begin{array}{c}
x_k y_k \\
\text{ellsworth}_k(x) \\
\text{arata}_k(x) \\
\text{see_on_tv}_f(x, x) \\
\text{shavonne}_k(y) \\
\end{array}
\]

In paraphrase: both women have a pure *de se* belief, i.e. one they’d express as “I will win”. But that is false in \( c_D \), Ellsworth would never express her own state of mind in this way; her belief, narrowly construed, is about the person she sees on TV. Because of the quantified \( z \)’s in \( R(z, z) \) there is no way to go back and pick a different relation from the context (besides, picking the relation of seeing someone on the TV would still give a false output). So if Zimmermann’s intuition is accurate, which I believe it is, we have a problem.

Before tinkering with the system, it’s important to see that we’re really not that far off the mark here. In fact, the truth conditions we predict are defended by Chierchia (1989:10), who claims that “[t]he most plausible interpretation of [a universally quantified report like (156)] claims that each person […] has a certain [pure] *de se* attitude” (full quote on page 133).
then, Chierchia probably didn’t consider mixed *de re/de se* contexts that would make his preferred reading false. Still, we might construe his remark as supporting our analysis, but only as a default strategy, i.e. it is what would happen in the absence of contextual pressure against it. What we’re looking for is an extension of the pragmatics of Acquaintance Resolution that allows us to derive a weaker meaning, true in mixed scenarios, while preserving the universally pure *de se* reading of Chierchia as the default. Evidently, we shall not resort to syntactic ambiguity.

Let’s think about the output that we want to get for (156). We have already seen a good If in 2.4.2.1, viz. (84), p. 134, the If that is predicted by simple (non-contextualized) relational and characterial analyses of *de re/de se*. In words it comes down to saying that for each of the two women there exists an acquaintance relation under which she has the proper relational attitude. Following 3.2.3.3 and 3.4.1 we can easily translate that If to an LDRT output context:

How to derive this from a context plus a PrelDRS? My proposal, first laid out in (Maier 2005b), is quite simple really; just think of $R$ as a presupposition that can be bound, yielding the Chierchian output of (162), or accommodated, thus introducing an $R$ in an appropriate universe, as in (163). So, what exactly do the presupposition and consequently the PrelDRS look like? How to resolve a second-order presupposition? If (163) is local accommodation, then how about global or even intermediate accommodation (and binding) options? As to the make-up of the presupposition: the universe must contain the second-order variable $R$, and since the contextual binding/unification of $R$ in the current framework is constrained by the fact that $R$ should hold between believer and *res*, we will put that as the presuppo-
3.4 Belief reports in LDRT

The rest of PrelDRS construction is now self-evident, as illustrated by the following example PrelDRS for the case at hand:

The next steps are as before, the context is added and the deepest presupposition, the pronoun, is resolved:

Then, we want to bind R. How to bind a second-order presupposition? By moving up the accessibility path, finding a place with appropriate material to equate the presupposed condition to, and then use that equation to resolve the presupposed variable through higher-order unification. If we add the preference for binding R to equality, we see that this presupposed acquaintance
proposal is indeed a conservative extension of what we were doing before. Let’s check that this is indeed the case by exploring the binding option in the current example. Local binding is preferred over global, and though there is nothing explicitly given in the local context, the default option of equality is available throughout the DRS, so we bind to that, i.e. we substitute equality for $R$ and get rid of the now superfluous presupposition box:

\[
\begin{array}{c}
x_k y_k \\
e lswordh_k(x) \\
arata_k(x) \\
see_{on}tv_r(x,x) \\
shavonne_k(y)
\end{array}
\begin{array}{c}
z_{fr} \\
woman_{fr}(z)
\end{array}
\begin{array}{c}
both_k \\
BEL_{z_{fr}} \\
u_k v_{fr} \\
center_k(u) \\
u = v \\
win_{fr}(v)
\end{array}
\]

The result is the strong Chierchian reading, exactly as before.

Now it turns out this is false, so we want to try a different resolution. Let’s consider our options. (i) Binding in the restrictor: won’t make a difference, the only candidate is still equality. (ii) Binding globally: that’s where we do find instantiations of proper acquaintance relations other than equality, but our presupposition cannot reach them because that would ‘unbind’ the $z$’s in its condition: a classic example of trapping (cf. 3.2.2), which not only prevents global binding but global accommodation as well. The next option (iv) would be the much debated intermediate or restrictor accommodation. Now, we can assume that second-order accommodation proceeds, technically speaking, exactly as first-order: just merge in the presupposed DRS at the locus of accommodation. Just note that this creates second-order DRSs, which are however no more exotic than second-order predicate logic. Also, since the acquaintance presupposition is unlabeled we can basically choose the layer that gives the best output, which is probably $fr$ here. The intermediate accommodation output looks as follows:
This is a somewhat hard to interpret DRS, because of the unquantified R appearing both in the restrictor and in the scope. It is however well-formed according to our definitions of duplex conditions with selective binding, as discussed in 3.1.4.1. The meaning we assign to this output can be translated to the more familiar generalized quantifier idiom as:

\[
(168) \quad \text{both}\left(\lambda z \exists R[R(z, z) \wedge \text{woman}(z)]\right)
\]

\[
(\lambda z \exists R[R(z, z) \wedge \text{woman}(z) \wedge \text{BEL}\exists u \exists v[\text{center}(u) \wedge R(u, v) \wedge \text{win}(v)]]\right)
\]

And in words as: the two women, who are somehow self-acquainted, each believe under some acquaintance relation that they’ll win. Since everybody, we said is acquainted to themselves under equality, the restriction that they are somehow self-acquainted is vacuous. This means that we have just arrived at the desired weak reading: for both women there is an acquaintance relation such that they believe they’ll win under that relation. Admittedly, the LDRT representation is not very pretty, and if it weren’t for some particular decisions in the semantics of duplex conditions earlier on, we might well have said it means something else or nothing at all. On top of that, many dismiss the intermediate accommodation option tout court, so if this was the final analysis I might lose some of my readers for relying on dubious, perhaps even arbitrary assumptions. Fortunately, there’s a final resolution option (v), local accommodation, yielding:
Exactly the output we wanted! And completely independent of peripheral, theory-laden decisions about unselective binding and intermediate accommodation.

This concludes my story about de re/de se reports under quantification. I’ve extended the Acquaintance Resolution account by making the acquaintance relation into a second-order presupposition-like entity that can project out of embeddings and either bind (by higher-order unification with a preference for equality) or accommodate. The resulting extension of Acquaintance Resolution is conservative in that in most cases it will behave exactly the same—the old system corresponds to the generally preferred binding option. We’ve considered in considerable detail a quantified example due to Zimmermann where this standard binding failed, and where local accommodation of the acquaintance relation gave exactly the right truth-conditions. The example is easily extended to other universal quantifiers, and to conservative quantifiers in general, but there’s reason to believe that only might offer some surprises, so we turn to that next.

### 3.4.3.2 Only

As previously discussed in 2.4.2.2, Percus & Sauerland (2003a) defend the Chierchian de re/de se ambiguity thesis. For this, they put an only quantified report in Zimmermann’s mixed scenario D, cf. (154). Their critical judgment has it that (170) has a true reading:

(170) Only Shavonne thinks she will win $\quad [=(85), \text{ p. 135}]$

They interpret this result as evidence that reports with co-referential pronouns must have a pure de se LF so that embedding that LF under only
suffices to derive the observed reading ‘Only Shavonne believes purely de se that she will win’. We’ve also seen how classical relational analyses assign the wrong reading, viz. only for Shavonne we can find some acquaintance relation (or characterial mode of presentation if you will) under which she holds the belief that she is winning. Contextualizing acquaintance seemed promising at first, because it could perhaps generate a reading that only Shavonne has the belief under her optimally salient acquaintance relation, which would be equality, presumably. Now that we have developed a proper dynamic framework for the resolution of acquaintance relations, let’s stop and see if this is indeed what we get:

(171) \[ \phi_D \sim \]

To refresh your memory of our particular only-quantifier, this is read as ‘Only for those z’s that are equal to Shavonne (s) does it hold that z believes the person she’s R-acquainted with will win.’ Now we resolve the presuppositions.

As for the pronoun, there’s two potential binding sites that match its content, the global y and the quantified z (which is semantically equated to y thus inheriting its semantic content).\(^{55}\) In such a case, the binding constraint LOCAL enforces a preference for the most local option:

\(^{55}\)Obviously, we assume that Ellsworth, x, is at this point unavailable—at least for the unstressed anaphoric pronoun she.
The choice between local \((z)\) and global \((y)\) resolution brings to mind our earlier analysis of the sloppy/strict ambiguity with pronouns embedded under \textit{only} in 2.4.3.4 (i.e. \textit{Only John did his homework} meaning (i) nobody else did theirs, or (ii) nobody else did John’s). The narrow resolution above would correspond to a sloppy reading, ‘Shavonne is the only member of the set of people who believe they’ll win’, while the global option would yield a strict reading, ‘Shavonne is the only member of the set of people that believe of Shavonne that she’ll win’. It turns out that the way we’ve set up the semantics of \textit{only} and/or Acquaintance Resolution ultimately collapses the strict reading into the sloppy, as the reader can easily verify, causing a slight concern over how to still derive the (admittedly marginal) strict reading in quantified coreferential \textit{de re} belief. Further research needs to be done, but for now we’ll just go ahead with the narrow resolution.

Time to resolve the acquaintance relation. Note there are no \(z\)-s in the main DRS, so we can rule out global resolution by trapping. Our 2nd order resolution strategy always checks the possibility of binding to equality first, and in this case that already works, in fact, it’s the only option, since there’s no other information available about the \(z\)-s:
3.4 Belief reports in LDRT

This is the desired, true output: the only person with a pure *de se* ‘I’ll win’ belief is Shavonne. Acquaintance Resolution’s prediction thus patterns with that expected of a contextualized relational analysis, though on second thought the exact derivation deviates from the intuitive idea of picking the most salient relation ‘from the context’, due to the trapping effect mentioned above. Below we’ll see the consequences of this discrepancy when we consider a final, crucial piece of data.

In 2.4.2.2 we went on to discuss the Zeevat example, meant to bring out a clear asymmetry between impure and pure *de se*. Zeevat’s observation is that, although we can truly say of the unique pure *de se* participant Shavonne that she is the only one who ‘believes she’ll win’, this is not true of the (equally unique) impure candidate, Ellsworth. In other words, in mixed scene D (174) is undeniably false:

\[
\begin{align*}
\text{(174) } & \# \text{Only Ellsworth thinks she’ll win} \quad [=\text{(90), p. 138}] \\
\end{align*}
\]

Following the same reasoning as above, we’d let *only* quantify over *de re* beliefs under a fixed acquaintance relation determined from the context. Since Ellsworth only has an impure *de re* belief about herself, we must take the acquaintance relation to be ‘seeing some Arata on TV’. Thus, we predict that (174) means *the only person who believes that the ‘Arata’ she’s seeing on TV will win is Ellsworth*, which, counter to all judgments, would be true in D. Chierchians might well construe this as a major argument against contextualized acquaintance, and thus, combined with the Percus and Sauerland argument against narrow scope relational analyses, against the unification of *de re/de se*. Moreover, it’s easy to see that a straightforward separatist analysis based on the ambiguity thesis does work: (174) comes out false on both the general *de re* (‘Ellsworth is the only one who has any type of *de re/de se* win-belief about herself’) and the pure *de se* LF (‘Ellsworth is the

<table>
<thead>
<tr>
<th>$x_k \ y_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{ellsworth}_k(x)$</td>
</tr>
<tr>
<td>$\text{arata}_k(x)$</td>
</tr>
<tr>
<td>$\text{see}<em>k \text{on}</em>\text{tv}_k(x, x)$</td>
</tr>
<tr>
<td>$\text{shavonne}_k(y)$</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{(173) } & \sim [R \Rightarrow] \sim \\
\end{align*}
\]
only one with a pure *de se* belief in her own victory’).

High time to investigate if this line of argumentation holds against Acquaintance Resolution as we’ve been developing it:

The default would be to bind $R$ to equality, but that gives a false reading:
This is as it should be, but we still have to check whether other resolution option might not make a true reading. What are the possibilities? Because the z's in the acquaintance presupposition trap R inside the quantifier, we only have to consider local and intermediate resolutions, and then only accommodation, since there's nothing to bind to except equality. Intermediate accommodation leads to the problematic only construction with an unquantified discourse referent in the restrictor. We've discussed this form in some detail in 3.1.4.2 where, lacking any clear clues as to an intuitively plausible interpretation for it, we eventually decided to simply ignore them. Here, under intermediate accommodation, we'd have a case of such a configuration arising. Blindly applying the technical definition would trivialize the restrictor's quantified acquaintance relation collapsing the intermediate output into the local accommodation truth conditions, exactly as happened with the universal quantifier in (167) above. This local accommodation then seems the only option:
Or: ‘Ellsworth is the only person for which there is a possibly non-egocentric acquaintance relation under which that person holds the belief that he or she will win’, which is false in D, since Shavonne also believes she, Shavonne, will win.

To sum up, Acquaintance Resolution beats Chierchian syntactic ambiguity analyses, first methodologically, replacing a syntactic ambiguity with a unified account of \textit{de re/de se} belief reports (discussed in 3.4.2), and second, empirically, as brought out by the quantified reports of mixed scenarios in the current section. More specifically, allowing the acquaintance relation to be accommodated like any other, first-order presupposition, as proposed in this section, properly accounts for the ‘narrow scope’ reading observed when reporting Zimmermann’s mixed scene reported with a universal quantifier, while preserving the general preference for pure \textit{de se} readings in the form of a resolution preference for binding to equality whenever possible (i.e. in coreferential reports). The examples with \textit{only} then showed that our system not only beats these separatist alternatives (countering explicitly the pro-ambiguity argument of Percus & Sauerland (2003a)), but also the original relational analysis and a straightforward contextualization thereof à la Abusch/Aloni. The last problem we have to face is that of the apparent grammaticalization of pure \textit{de se} reporting, which was the original motivation for Chierchia’s proposal, and the related data involving monstrously shifted first person indexicals.

### 3.4.4 Pure \textit{de se} reports and shifted indexicals

Our argument in favor of a unified account of \textit{de re/de se} is almost complete, but we’ve saved the best counterarguments and data for last. We claim that in belief report semantics, we should generate (compositionally) but one logical form, one preliminary DRS that is, which is underspecified for \textit{de re/de se}, and then a general pragmasemantic resolution module will determine whether, given this PrelDRS and a context, the output is going to be pure \textit{de se}, or \textit{de re} under some particular relation of acquaintance. But then, what to do with examples like (178) that are commonly taken to have only pure \textit{de se} truth conditions?

\begin{equation*}
\text{(178) Ellsworth believes to be on the winning side}
\end{equation*}

\begin{equation*}
\text{[=(61e), p. 220, =(1c), p. 72]}
\end{equation*}

At this point, one may be tempted to cop out and adopt a kind of ‘hybrid analysis’, i.e. an unified, contextualized \textit{de re/de se} analysis of coreferential pronoun reports built on Acquaintance Resolution, which assigns a separate
pure \textit{de se} DRS to report construction with infinitival and gerundial complements. In fact, this is exactly what was suggested in 3.2.3.3, see in particular (76), p. 233. As an added bonus, perhaps, we could probably implement this without having to postulate a silent PRO element that fills the complement’s subject role. For sure, if this was all there was to the data, I’d be the first to adopt this strategy.

Unfortunately, there’s more. First, there are logophoric pronouns, which, glossing over some details and reservations spelled out in 2.4.1, also have the effect of making a report exclusively pure \textit{de se} (cf. Kusumoto’s (1998) (55), p. 116), but now with a surface form that structurally matches the normal \textit{de re}/\textit{de se} ambiguous embedded pronoun construction, i.e. with a clearly realized pronoun (glossed \textit{LOG}) as embedded subject:

\begin{align*}
(179) & \quad \text{kofi be \text{\`e} dzo} \\
& \text{Kofi say LOG -leave} \\
& \text{‘Kofi said that he left’ (Kofi: “I leave/left”)} \quad \text{[Ewe, =\text{(54)}, p. 116]}
\end{align*}

To fit this into the hybrid account we’d have to assume that, despite appearances, a logophoric pronoun is something rather different than a third person anaphoric pronoun; the logophor is not treated as a presupposition trigger but rather as a marker that tells us to construe the whole complement as purely \textit{de se}. Still, this can be, and has been, done, e.g. along the lines of Chierchia’s proposal which actually needs some element (so, PRO or LOG) to serve as a \textit{\&}-abstractor/variable binder for deriving pure \textit{de se} truth conditions.

But there’s still more. The final argument against the kinds of unificatory and hybrid approaches under discussion, which is due to Schlenker (1999,2003), involves shifted indexicals and was discussed in detail in 2.4.3. We focus on clear cases where a morphological first person, embedded under a third person (attitude) reporting verb, is interpreted as a pure \textit{de se} pronoun:

\begin{align*}
(180) & \quad \text{jon j\text{\^o}gna n\text{o}-\text{n}\text{n} yil -all} \\
& \text{John hero be -1.sg say.3.sg -aux.3.sg} \\
& \text{‘John says that he is a hero’} \quad \text{[Amharic, =\text{(50)}, p. 111, =\text{(93)}, p. 141]}
\end{align*}

This is just the regular Amharic first person pronoun, so, as argued in 2.4.3.1, this does not fit the hybrid analysis originally proposed for infinitival reports. There’s a clearly realized pronoun, an indexical even, in the complement and I take it that any proper account of these data will have to treat it as such, as a \textit{k}-presupposition in our current framework.
This will be my starting point, the Amharic first person, embedded or not, is a real indexical and as such it triggers the corresponding \( k \)-presupposition. If we get that working, it seems reasonable to extend the same mechanisms to treat LOG and PRO as pronominal elements, similar to Amharic \( I \) (though, as we shall see, each with its own particular resolution restrictions). In this way I hope to do justice to my overall objective, which is to generate PrelDRSs straightforwardly from the surface structure, and take the load off syntax to do the work more effectively in the semantics-pragmatics interface, but some concessions will have to be made in this area, on any account. What I need to postulate is, first, a silent PRO for gerunds and infinitivals, which functions rather like a first person indexical pronoun. Second, and this is perhaps one of weakest points, I must assume that logophors too are semantically first person. I will return to this point in 3.4.4.4 below to suggest an alternative account that fixes this, at the cost of losing the overall unification of the three ways of conveying unambiguously pure \( de \ se \) (PRO, LOG, shifted \( I \)). Now we turn to the shifted first person.

3.4.4.1 Acquaintance with concepts?

To bring out the problems caused by shifted indexicality, let’s just see what happens when we apply Acquaintance Resolution to a pseudo-Amharic report with a shifted first person, treated as a regular indexical presupposition. For uniformity I’ll switch from the well-known Amharic speech report to a pseudo-Amharic gloss of a belief report. With some minor modification the entire reasoning below will go through for speech reports, but we’ll return to this matter in 3.4.4.2, where we also briefly consider other attitudes. Still, I apologize for initially manipulating the data somewhat, especially since it is as yet unclear what types of reporting verbs allow shifting in which languages.\(^{56}\)

\[ \text{(181)} \quad \text{John, thinks I, am a hero} \quad \text{[pseudo-Amharic, } \approx \text{(180)}] \]

---

\(^{56}\)With logophors it has been noted that there seems to be language variation according to a hierarchy of reporting verbs, cf. 2.4.1.2.
Assuming as always that the global context provides a John, and a speaker, viz. the one who utters the report, this PrelDRS will only give us non-monstrous de re readings:

If the context would further provide some kind of acquaintance link between John and the actual speaker, me, we could bind $R$ and finish the derivation to get a nice de re output that reports John’s opinion of me, the utterer of (181), rather than about himself.

Now, if we’d been dealing with English this would be just right and completely in line with the Kaplanian prohibition on monsters, which posits that indexicals are never shifted. Moreover, even the Amharic analogue can be read in this way, as a report of what John thinks about me, so that’s something at least. However, crucially, in languages like Amharic, Zazaki or Ancient Greek, a second, pure de se, shifted, reading is observed. We can easily represent this by now familiar desired output:
Chapter 3. A semantics of attitude reports

The question becomes how do to derive this output. Obviously the first resolution step in (182) was already wrong, for I am quite distinct from John, which obviously blocks the possibility of binding the acquaintance relation to equality. We must find a way to bind the first person presupposition to John, but he’s a third person. The solution is to bind it to the embedded attitude center \( u \), which is first person and which is not quite John, but a pretty good approximation of him, for it’s the center of his belief set, the person he believes himself to be.

The first step then is to modify \( \textbf{Prel} \) so as to output a preliminary structure where the embedded subject’s presupposition is triggered within the belief box. The easiest way to do this is to detach the presupposition providing \( w \) from the free variable \( w \) in (181)’s \( R(z,w) \) and put it inside the belief box. At least then it seems that \( u \) is accessible to it, so we can proceed as follows:
3.4 Belief reports in LDRT

(185) \[ \begin{array}{c|c}
  x_k y_k \\
  \hline
  \text{john}_k(x) \\
  \text{speaker}_k(y)
\end{array} \] \[ \leadsto \ominus (184) \leadsto \begin{bmatrix}
  z \mapsto x \\
  w \mapsto u
\end{bmatrix} \leadsto x_k y_k \]

But this is a dead end too, for with a free variable \( u \) in the body of the acquaintance presupposition we’ll never be able to bind or accommodate a suitable acquaintance relation.\(^57\) Intuitively though, we’re not that far off: if we think of \( u \) as \( x \)'s self-image, then what we want to bind \( R \) to is ‘the person you believe to be’, the relation that holds between a person and her belief center. To approximate this I follow a suggestion of Zimmermann (p.c.) to allow individuals to be acquainted to (a certain restricted class of) intensional objects, i.e. individual concepts. For John, one such concept could be John’s self-concept, i.e. the function that maps each of John’s belief alternatives to its center. The advantage of the self-concept over the extensional \( u \) is that the concept can occur representationally outside the scope of John’s belief DRS, which is where we’ll let Prel generate it. One last change concerns the presupposition triggered by the embedded subject, which has yet to be linked to this underspecified concept. The choice for linking the presupposed referent to the extension of the concept will become clear in the course of the computations below. Now, concretely:

\(^57\) Actually, the free variable does not stop Maier (2005a) from proposing a somewhat simpler, extensional alternative in which the free variable does not prevent singling out a similarly open piece of context DRS to unify with, essentially assuming a purely formal, symbolic notion of unification. An important difference between that proposal and the current analysis is the fact that the former cannot be extended to incorporate the presuppositional (=semantic) analysis of acquaintance needed for quantified reports.
We see that $I, w$, is happily bound to the belief center $u$, without creating an open DRS. Moreover, although we have given up on $w$ being the object of acquaintance directly, we do retain a rather close connection between them, given by the $w = \forall c$ clause in the second presupposition. It means that, wherever $w$ ends up, it gives the extension of the concept the subject is acquainted with. Note here already that if the embedded presupposition had floated up to top-level, it would restrict the acquainted concept to one that actually denotes the current center, i.e. me. We return to this after we having finished the monstrous reading, in order to ensure that the whole intensional machinery proposed here preserves the earlier results.
3.4 Belief reports in LDRT

Back to the derivation, first let’s simplify our DRS by getting rid of the u (obsoleted by the equation with $\lor c$):

$$
\begin{array}{|c|c|}
\hline
x_k & y_k & c \\
\hline
john_k(x) & speaker_k(y) & R(x, c) \\
\hline
\end{array}
$$

Then bind R. Obviously equality doesn’t fit anymore since x is an individual and c is a concept. The strongest analogue of equality in such a case would be necessary equality, i.e. when c denotes a rigid concept and x corefers with its constant extension. But this does not apply here either, so we start searching through the DRS for a suitable relation between x and c. To further narrow down the search note that we’re looking for something given by the input context already, not something newly contributed by the report. Strictly speaking there is nothing suitable in our input, but part of what we added with our new PrelDRS may safely be assumed given, and that’s exactly the part that contains both the x and the c: the belief of John minus the acquaintance relation and the complement’s asserted contribution (being a hero).

$$
\begin{array}{|c|c|}
\hline
d & center_k(\lor c) & R(\lor c, d) & hero_{fr}(\lor d) \\
\hline
\end{array}
$$

What kind of acquaintance relation is this? Well, it’s the relation that holds between an individual and a concept if that concept’s extension in the individual’s belief alternative denotes the center of that belief alternative, i.e. iff the concept is the individual’s self-concept! Continuing the derivation by binding R with the above substitution:
Remember, we’re trying to show that Acquaintance Resolution, with the intensional twist and the addition of the global concept, \( c \), is going to output the pure \( \text{de se} \) reading represented in (183). In the next subsection I’ll finish the argument by showing how a well-known principle of belief logic can be applied to derive (183) from (190).

### 3.4.4.2 Introspection: a rationality criterion

The thing to note in (190) is that there’s a belief-operator embedded under another belief operator. For those familiar with the analysis of belief in modal logic, this may ring a bell and remind them of the principles of introspection. For a while, the aim of modal logic was to provide axiomatic characterizations of different logics for different modalities (belief, necessity, knowledge, etc.) and prove consistency and other logical properties of these systems. The most famous ones are perhaps the minimal modal logic \( K \), the basis of all others, and \( S5 \), the logic commonly used for modeling knowledge and metaphysical necessity. Belief is often modeled as \( S5 \) minus the principle of veridicality, \( \Box \varphi \rightarrow \varphi \), which for the modality belief would amount to the claim that if \( \varphi \) is believed, it is true. What remains are the following axioms and rule:

\[
\begin{align*}
\text{(191) a.} & \quad \text{rule: if } \vdash \varphi, \text{ then } \vdash \Box \varphi \\
\text{b. axioms:} & \\
K & \Box[\varphi \rightarrow \psi] \rightarrow [\Box \varphi \rightarrow \Box \psi] \text{ (distribution)} \\
D & \Box \varphi \rightarrow \neg \Box \neg \varphi \text{ (consistency)} \\
(4) & \Box \varphi \rightarrow \Box \Box \varphi \text{ (positive introspection)} \\
(5) & \neg \Box \varphi \rightarrow \Box \neg \Box \varphi \text{ (negative introspection)} 
\end{align*}
\]

Adding these three axioms to \( K \) gives a reasonable logic of belief, one where a person’s beliefs are closed under logical consequence, she believes all tautologies, has internally consistent beliefs, and is aware of the things she believes and the things she does not believe. The closure property amounts to logical
omniscience, but, as we discussed in 1.1.2, that simply cannot be helped in modal logic. The consistency criterion seems reasonable enough, as does positive introspection. Some may feel that negative introspection is too strong, that there may be propositions we do not believe without actually believing that we do not believe them. Note then that we're looking at a rather weak notion of belief, as shown by the omniscience predictions discussed in chapter 1. In any case, let's just go along and add it, for the sake of argument. The resulting logic is called \textbf{KD45} and it's not hard to prove the following:

\begin{equation}
D, (5) \vdash \Box \Box \varphi \rightarrow \Box \varphi
\end{equation}

Sketch of a proper derivation: First, by transposition and double negation cancellation we see that axiom (5) is equivalent to \(\neg \Box \neg \Box \varphi \rightarrow \Box \varphi\). Now, start with \(\Box \Box \varphi\). Modus Ponens with axiom D gives \(\neg \Box \neg \Box \varphi\), which, with the new form of (5), yields \(\Box \varphi\). QED.

So, if you believe that you believe something, you believe it. And the other way around, as given by (4). Considering that our intensional systems always capture a somewhat idealized notion of belief, this is not an implausible principle at all, and, apparently, it is valid in the widely used belief model \textbf{KD45}. If we could formulate and add an analogue of this ‘reverse introspection’ principle to our DRT semantics of \textit{bel} we could reduce the double embedding of (190) and, hopefully, get a step closer to the desired output, (183). There are two ways to go about this, syntactic or semantic. The first means that we formulate DRS conversion rules or axioms to capture the idea of reverse introspection. The second is to translate the syntactic property to a property of the underlying models, or rather, of the accessibility relations of the models’ interpretation of the modality. Let’s consider the syntactic option.\textsuperscript{58}

The representational, procedural view of interpretation in our formulation of LDRT+PA, lends itself very well to the addition of symbolic operations inspired by traditional axioms. The idea would be to add to the resolution process a subprocedure for simplifying complex belief embedded DRSs, based on the classical introspection principles that say that multiple belief embedding is equivalent to single belief embedding. Such procedures are easily formulated, the only things we have to look at still are the subject parameter \(x\) in \textit{Bel}_x and the fact that we are dealing with belief alternative contexts with centers, rather than plain old worlds.

\textsuperscript{58}As for the semantic option, we could enforce our reverse introspection principle (plus positive introspection) by restricting attention to models where, if \(c \in \textit{Bel}(a, w, t)\) and \(b\) is the center of \(c\), then \(\textit{Bel}(a, w, t) = \textit{Bel}(b, w_c, t_c)\) (the belief set of the center of one of your belief alternative contexts is the same as your belief set).
In classical doxastic predicate logic, the subject who has the belief is often treated by parameterizing the modal operator with a term denoting that individual, which is how we got our notation \( \text{BEL}_x \). What this means is we have a whole bunch of distinct belief operators, one per individual, each adhering to the axioms in (191). Reverse introspection, (192), then amounts to the reducibility of two or more belief operators with the same subject parameter in a row: \( \text{BEL}_x \text{BEL}_x \varphi \rightarrow \text{BEL}_x \varphi \). Obviously we cannot similarly reduce a statement expressing that \( x \) believes that \( y \) believes such and such. But if we look at (190) again, we see that the indices denoting the believers are distinct. Fortunately, the embedded believer \( \vee c \) is not really that distinct from the first, \( x \), since \( \vee c \) denotes the center of \( x \)'s belief alternatives. If you think about it that seems exactly what's needed, given the context/property framework adopted since 1.2.3. We should adapt the classical \( \text{BEL}_x \text{BEL}_x \varphi \rightarrow \text{BEL}_x \varphi \) (“if \( x \) believes that \( x \) believes that . . . ”) to something like \( \text{BEL}_x \lambda u[\text{BEL}_u \lambda u'[\varphi]] \rightarrow \text{BEL}_x \varphi \) (“if \( x \) self-ascribes the property of self-ascribing . . . ” or “if \( x \) believes that he* believes that . . . ”). Actually, I am not up to the task of properly reformulating the whole of KD45 into a proper axiomatic proof system for modal predicate logic that takes essential indexicality into account, so we'll just concentrate on reverse introspection, and on DRT. I propose the following introspective simplification rule for (L)DRT:

\[
\begin{array}{c}
\text{BEL}_\xi \\
\vdots \\
\text{BEL}_\eta \varphi \\
\vdots \\
\vdots \\
\text{BEL}_\xi \\
\end{array}
\sim
\begin{array}{c}
\text{BEL}_\xi \\
\vdots \\
\text{center}(\eta) \\
\vdots \\
\vdots \\
\vdots \\
\end{array}
\]

It should be evident how such a schema is to be read and applied to double belief embeddings. Furthermore I take it that the above discussion has provided ample justification for adding such a principle, which adequately captures the idea that believing to believe something means that you believe it. Applying it to (190) will make it even more clear. To do so, we take \( x \) for \( \xi \), \( \vee c \) for \( \eta \) and the most deeply embedded box for \( \varphi \), and we get:
On the further, rather uncontroversial, assumption that beliefs can have but one center, we conclude that $\forall c = \forall d$. Now notice how the intensional variables no longer play any role, so they might as well be replaced by the more familiar extensional ones. With this, we reach our destination, (183):

What we have shown so far is that going intensional plus a few other minor adaptations to Prel, and the addition of a general introspection principle of rational belief can give us the right output for Amharic, shifted $I$. The important selling point of my account is that it does not rely on feature deletion or de-rigidification of the indexical in the syntax. Everything is done in the pragmasemantic resolution module. The difference between the Amharic shifted and the English (or Amharic, for that matter) rigid $I$ is merely a matter of a resolution ambiguity of the same $k$-presupposition: English $I$, apparently, must be bound globally (and yes, we have to stipulate that in the lexicon somehow), while Amharic $I$ can also be bound by a local speaker or center in a $k$-layer. We’ll get back to this lexical difference in resolution behavior between Amharic and English when we also consider the other manifestations of pure de se reporting in 3.4.4.4.

One last point we will briefly discuss here concerns the other attitudes. We’ve been focusing solely on belief in this thesis, but the idea behind that was that the methods developed for de re and de se belief would naturally extend to the other attitudes (know, hope, fear, etc.) and to speech reporting.
However it is not immediately clear that the addition of an introspection rule is a sensible one for other attitudes than believe? Well, for the attitude of knowing it’s intuitively OK, and for wanting and hoping too, but what about doubt? “If I doubt that I doubt that $\varphi$ then I doubt that $\varphi$”, seems absolutely false, and the same goes for other ‘negative attitudes’, like fear and hate. It seems that we can only account for shifted indexicality (and, by extension, pure de se reporting, see 3.4.4.4) for ‘introspective attitudes’. The data show no evidence for this generalization whatsoever, in fact, one of the things the rather scarce data set does show is that shifted indexicality happily occurs at least in speech reports, and I’m not convinced that indirect speech reporting involves an introspective attitude—can one consistently say “I’m saying that I’m saying that John is coming, I am not saying that he’s coming”?

In any case, it seems unsatisfactory to restrict ourselves to introspective attitudes, especially since we’re going to apply the same machinery to PRO and LOG. There’s no denying that some negative attitudes allow pure de se reports with PRO constructions, even in English (e.g. I hate doing/to do this). For a way out, remember why we restricted ourselves to belief in the first place: the asymmetric dependency between belief and the other attitudes, which we described as necessarily rooted in an underlying belief base (cf. p. 35).

It turns out that this complication with attitudes other than belief may help us out of a jam here. The idea is as follows: Take a de re doubt, say, Shavonne doubting of Ellsworth that she’ll win. Applying our relational explication of de re belief means that an acquaintance relation between Shavonne and Ellsworth is presupposed and Shavonne doubts that the person she’s so acquainted with will win. The second requirement highlights the attitude’s dependence on belief, since it cannot mean that in all doubted worlds she is acquainted with a certain individual who then wins. Rather it’s the person she believes to be so-and-so acquainted with that she fears will win. The fear attitude only pertains to the new information, the predicate win; the acquaintance relation and for instance the attitude’s center ‘float up’ to a basic, implicit belief attitude. As I said before, I cannot offer any account of this yet. Perhaps LDRT could help, dividing an attitude embedded DRS into different layers—one for belief, to be evaluated in the belief alternative

\[59\] In the past tense, this sentence seems much more acceptable, especially when you stress the second said:

(i) I said that I said that John’s coming, I didn’t say he’s coming

Part of the reason for this however is that the saying events are assumed to take place at different times, which breaks the argument.
contexts, another for *hope*, evaluated in the hope alternatives, etc.—so we can at least adequately represent the interdependence of the different attitudes, which often share their discourse referents, a prime characteristic of LDRT.

In any case, assuming that in representations of *de re* attitudes everything except the part about winning actually falls under a belief operator, we don’t have to worry about the other attitudes anymore. Once we have ensured a proper belief-fear hybrid representation, we need only assume introspection for the belief part to properly reduce double embeddings of the type exemplified by DRSs like (190). The dependence on a basic belief layer, which is assumed to be introspective, thus helps us extend our theory of belief to the other attitudes, regardless of their logical properties.

3.4.4.3 Conservative?

Now that we have this nice Prel-extension, featuring acquaintance to concepts instead of individuals, we have to show that all the previous results (Ortcutt, the ambiguous *de re/de se* pronoun, English *I*, etc.) are retained. Let’s recompute one representative example report from before, but now in the extended intensional framework, to see what we come up against. Let’s just take one half of the Ortcutt example from 3.4.2.3:

(196) Janell thinks Ortcutt is dumb

\[\text{Janell thinks Ortcutt is dumb} \quad \text{[=(147), p. 286, =(70a), p. 227]}\]
Let’s reconsider the intuitive motivation behind our shift to intensional acquaintance. What does it mean to be acquainted with a concept? It seems a rather rare occurrence, what kind of concepts do you really know, intimately, preferably through the senses? The concept of yourself is one for sure, and that’s the one we happily made use of before, but apart from that it may be safe to assume that, if at all possible, we’ll be looking for acquaintance relations that relate us to individuals, or rigid concepts, if you wish to retain concepts formally. Time for the formulation of yet another violable constraint hidden in the definition of what it means to be a proper acquaintance relation: avoid truly intensional acquaintance relations by considering rigid concepts. Now a rigid concept, one that denotes the same individual in all possible worlds, may safely be identified with its extension, i.e. if \( c \) were rigid, we might as well replace it with its extension \( \vee c \), and the same for \( d \), while replacing the intensional \( R \) with its extensional counterpart \( R’ \). At that point, the intensional variables become redundant and a familiar, extensional DRS starts to appear:
3.4 Belief reports in LDRT

\[ x_k \, y_k \, \text{c} \]

\[
\begin{align*}
\text{janell}_k(x) \\
\text{ortcutt}_k(y) \text{famous}_p(y) \\
\text{admire}_p(x, y) \\
\text{write}_p(#042_p(y) \text{referee}_p(x, y) \\
\end{align*}
\]

\[ y = \text{c} \]

\[ \text{BEL}_{x_i} \]

\[
\begin{align*}
\text{center}_p(u) \\
\text{R'}(u, \text{c}) \\
\text{dumb}_p(\text{c}) \\
\end{align*}
\]

\[ R' \leadsto \lambda s \lambda t [\text{write}_p(#042(t) \text{referee}(s, t))] \]

\[ x_k \, y_k \]

\[
\begin{align*}
\text{janell}_k(x) \\
\text{ortcutt}_k(y) \text{famous}_p(y) \\
\text{admire}_p(x, y) \\
\text{write}_p(#042_p(y) \text{referee}_p(x, y) \\
\end{align*}
\]

\[ \text{BEL}_{x_i} \]

\[
\begin{align*}
\text{center}_p(u) \\
\text{R'}(u, v) \\
\text{dumb}_p(v) \\
\end{align*}
\]

[cf. (147), p. 286]
This is the exact same acquaintance resolution and output we had computed earlier on.

This example derivation concludes our demonstration of how the plain acquaintance resolution strategy can be embedded as a special, default case of the more general intensional system developed in this final section, viz. the case where the acquainted concept is assumed to be rigid and the relation extensional.

3.4.4.4 PRO and LOG again

Now, let’s extend our analysis of shifted indexicality to the remainder of the pure de se data, involving PRO and LOG reports. We have just seen how an embedded first person can be bound by a belief embedded center, leading, through a slightly modified mechanism of Acquaintance Resolution, to a pure de se reading. So, if we simply treat PRO and LOG as manifestations of a first person presupposition, i.e. like Amharic I, we’re almost there. Take for instance the PRO report about Ellsworth:

\[(199)\] Ellsworth believes PRO to be on the winning side \[\Rightarrow (178)\]

Unfortunately, we’ve already shown that such an LDRS has two possible outputs, corresponding to the pure de se (local binding of center presupposition) and the Kaplanian de re reading (binding to global speaker, i.e. Ellsworth thinks I, Emar, am winning). The second reading was desired for Amharic I (and Zazaki, Ancient Greek and others), and for English I, where it should really be the only possibility. Just as we simply stipulated that
English $I$ must take widest possible scope, so we can now stipulate that PRO and LOG need to be bound locally, or rather, as von Stechow (2002) shows with multiple belief embeddings, that PRO needs to be bound by the most local center, while a logophor allows intermediate bindings as well, as long as the global center is avoided.

The kind of stipulations we’re talking about here come from the lexicon, and need to be represented in the PrelDRS somehow, or be added as resolution constraints (a bit like Montagovian meaning postulates). The important thing to note is that they are stipulations, but there’s no way to get rid of them, because there clearly is a lexical difference between English $I$, Amharic $I$, LOG, and PRO, and this difference has to be encoded somewhere. Schlenker calls it the typology of indexicals, and he captures it with a so-called filtering mechanism that introduces the features $\pm\mathrm{ACTUAL}$ and $\pm\mathrm{C(ontextual)}$ (cf. p. 160). Our account differs in using resolution restrictions instead of the $\pm\mathrm{ACTUAL}$ feature, and using a $k$ label instead of $+C$ for marking indexicality, as illustrated in table 3.4 below. Another difference concerns the treatment of PRO and LOG’s features (are they 1st or 3rd person?), to which we turn next.

<table>
<thead>
<tr>
<th></th>
<th>Schlenker/von Stechow</th>
<th>Acquaintance Resolution</th>
</tr>
</thead>
</table>
| $I_{\text{Eng}}$ | 1.sg $+$C $+$ACTUAL   | $x_k$  
|                |                        | $\text{center}_k(x)$     |
| $I_{\text{Amh}}$ | 1.sg $+$C              | $x_k$  
|                |                        | $\text{center}_k(x)$     |
| LOG            | 3.sg $+$C $-$ACTUAL    | $x_k$  
|                |                        | $\text{center}_k(x)$     |
| PRO            | 3.sg $+$C $-$ACTUAL    | $x_k$  
|                |                        | $\text{center}_k(x)$     |
| he             | 3.sg $-$C              | $x$                             |
|                |                        | $3_{\text{sg.m}}(x)$         |

Table 3.4: Typology of indexicals/pronouns

What about the features of LOG and PRO? We let Prel assign them a first person representation merely because the semantics dictates that they
be bound by a local center. On the other hand, there is evidence that, on the surface, PRO and LOG are third person. Take (200):

(200) John hopes PRO to buy \{*myself
himself\} a car

\[=(135), p. 160\]

It’s clear that the reflexive has to be bound by the PRO element (commanding it), which suggests strongly that PRO has third person features, even if semantically it is first person. And for logophors it’s even clearer in some languages, because logophoricity can be marked on the verb clearly separated from some pronoun or morpheme denoting the embedded subject and marked as a third person (cf. also (57), (58), p. 117):

(201) \(a\`e\text{ ko } a\`e \text{ do } -\text{LOG}\)

\['\text{he said he fell}.’\]

[Gokana, (Hyman & Comrie 1981:20)]

For Schlenker and von Stechow-like systems there’s no problem; they can resort to an agreement mechanism to delete an agreeing third person feature. In (201), for example, *John* and PRO (as a third person) agree, so we delete the third person of PRO at LF and get a pure de se interpretation, just as we did when deriving the (preferred) pure de se interpretation of *Ellsworth thinks she’s going to win*. We, however, decided to get rid of agreement/feature deletion and interpret those third person pronouns as is, so for us this new discrepancy between the outwardly third person PRO and LOG, and their first person semantics is definitely a problem. And unfortunately a problem for which I do not yet have a satisfactory solution. But let’s go through a couple of half solutions, and one promising idea that deserves more attention.

For PRO, perhaps a possible reaction would be to deny that they are really third person: they are first person and the form of the reflexive in (200) needs an independent explanation. I simply don’t know if something along these lines can be done and if so, if it would extend to logophors, so this requires further research. A second possibility is to admit a limited form of agreement to delete these surfacy features, just like we relied on a limited agreement rule for quantifier binding (cf. 3.4.3.1). However, it remains to be seen whether such a restricted agreement rule makes sense, i.e. whether we can clearly demarcate just these cases so as not to be forced into implicitly adopting the full feature agreement/deletion system used by Schlenker and von Stechow.

As a final option, we could follow a more radical suggestion of Schlenker (p.c.), which is to see PRO and LOG constructions not as first person pronouns, but as the linguistic surface manifestations of the fact that the ac-
quaintance relation is set to equality. To see how this could work, first take a look at the realization of logophoricity in Eleme or Gokana, where the LOG feature is realized separate from a third person pronoun, as in (201). The current analysis as summed up in table 3.4 is committed to the claim that the LOG feature on the verb combined with the third person pronoun introduces a semantically first person presupposition. By contrast, we could also say that the embedded subject really is just a third person pronoun, as in the English translation, and that LOG restricts the resolution of acquaintance to equality. A PrelDRS of the Gokana example (201) would look something like this:

(202)  he said he fell-LOG

\[
\begin{array}{c}
\text{R(x,y)} \\
\text{3.sg.masc(x), 3.sg.masc(y),} \\
\text{fall(v)}
\end{array}
\]

This immediately accounts for the pure \emph{de se} character of LOG reports, as opposed to their natural English translations, which of course allow both pure \emph{de se} and other, contextually specified \emph{de re} readings. Note also how this new type of analysis obviates the intensionality hacks and the use of introspection axioms described in the previous subsections. However, since I don’t see how to extend this new idea to account for Amharic \textit{I} in a sensible way, those mechanisms still remain in place for handling shifted indexicals.

From Gokana’s 3.sg+verb+LOG-complements we can extend the new analysis to logophoric pronouns, as in Ewe (cf. (54), p. 116), which are then just special forms that introduce both a third person pronoun presupposition plus the requirement that the acquaintance relation is equality. And exactly the same holds for PRO.\footnote{In fact, if you know beforehand that the acquaintance relation is going to be equality, there’s really no need to resolve the embedded pronoun, so we could leave the whole pronominal aspect out of the representation of PRO and Ewe-style LOG. A reason to keep the presuppositional representation of the third person \textit{res} is for uniformity with...}
promising line of future research to determine if it will really solve our issues with LOG and PRO within the Acquaintance Resolution framework.

the Gokana-style complements, and perhaps to account for the third person reflexives in reports like (200).
Chapter 4

Conclusions and discussion

Through this thesis I have argued that the *de se* mode of believing and belief reporting is not as special as it is often made out to be. I have shown that pure and other *de se* beliefs and reports are best analyzed as instances of relational *de re* belief. Combining this reduction of *de se* with the contextual resolution of acquaintance relations eventually led to a new semantics of attitude reports. To recapitulate how we arrived at that point, let me highlight the main contributions of each chapter. After that, I’ll point out some issues left open for future research.

4.1 Conclusions

Besides laying the theoretical, methodological and terminological foundations for this thesis, chapter 1’s main contribution was the discovery that, under standard definitions, *de re* (=beliefs about actual objects) and *de se* (=beliefs the subject could not have expressed without indexicals) characterize more or less the same set of beliefs. So-called pure *de se* beliefs (about myself, here, now) then form a special, proper subclass thereof. The unified analysis of *de re/de se* as relational attitudes encountered in 1.2.4.1, a synthesis of Kaplan’s ‘vivid names’ with Lewis’ property self-ascription, accommodates this rather well. A relational belief that \(x\) is \(P\) means that the believer is somehow acquainted with the *res* \(x\) and that he or she self-ascribes the property of being so acquainted with someone \(P\). Different kinds of *de re* and *de se* beliefs simply correlate with different kinds of acquaintance relations. This is illustrated in table 1.1, p. 54, which shows among other things how pure first person *de se* belief corresponds to the acquaintance relation of equality, which receives a special status when we turn to attitude reports.

Comparing this relational account with its most popular and equally uni-
fied rival analysis, Kaplan's two-dimensional characterial belief semantics, reveals that both analyze *de re* belief as belief 'under a suitable description', and it's the nature of that description that makes it a particular kind of *de re*, like, say, pure *de se* belief. Also, the belief objects in both are more fine-grained than classical propositions (sets of possible worlds): both eschew full sententialism but they do use a notion of context, modeled as a situated possible world, to incorporate point of view. In fact, by restricting attention to the diagonal of a character it's easily seen that both involve properties being self-ascribed. The difference between believing a character and believing a proposition under a certain acquaintance relation becomes more apparent once we transpose the belief semantics to a report semantics in the next chapter.

In chapter 2 we looked at the semantics of *de re* and *de se* reports. The relational semantics advocated in the first chapter works quite well when interpreted as a semantics of reports, but it's not flawless. A first minor issue has to do with compositionality, but we concluded that that could not be helped: we need to postulate a kind of movement that separates the *res* from the content of the ascription. In line with this observation, note that the popular compositional alternative, Kaplan's character-based analysis of indirect discourse and belief reports, was recently shown to be entirely inadequate as a semantics of belief reports.

One very promising extension of the relational unified framework was the pragmatic determination of acquaintance relations in the context. But even that couldn't help us with the main problem faced by unified *de re/de se* analyses: the existence of unambiguously pure *de se* reports (e.g. complements with infinitives and logophors). Other arguments against unified *de re/de se* semantics (contextualized or not) discussed in 2.4 involve belief reports embedded under *only*, and monstrously shifted indexicals. The analyses that have been proposed so far to account for these kinds of data all introduce some form of syntactic ambiguity (Chierchia's ambiguity thesis) into *de re/de se* reports and/or an intricate morphosyntactic account of feature deletion under agreement (Heim, Schlenker, von Stechow). One of the main motivations behind this work was to do away with all this morphosyntactic theory in favor of a more transparent pragmatic/semantic analysis. Thus, my aim for chapter 3 is to further develop the contextualization idea for the relational attitude analysis into a general, pragmatic/semantic account that can deal with all the data considered so far.

Chapter 3 presented the dynamic framework needed for my contextualized version of *de re/de se* as relational attitudes. The move from static to dynamic semantics was inspired by the need for a sound and more subtle way of determining acquaintance relations in a given context. It seemed that the
representational framework of DRT with the powerful theory of presupposition as anaphora (PA) would fit the bill, so we defined the basic notions that we’d be needing to accommodate a semantics for our reports (e.g. preliminary structures, intensional DRT semantics and duplex conditions). I went beyond standard DRT+PA in adding layers, primarily to give a principled account of referential terms such as indexicals and proper names. The idea was that by separating rigid and ‘at issue’ content at different layers we can give a proper two-dimensional semantics without relying on external anchors, which I argue to be ad hoc devices, alien to the representational nature of DRT. Combined with the anaphoric theory of presupposition, this layering led to a distinction between rigid and descriptive presuppositions. This way we were able to account for some recent data involving non-globally bound and accommodated proper names. My main motivation for doing all this here however has to do with indexicals, which I need to be both rigid and bindable in order to account for their shifty behavior in e.g. Amharic and Ancient Greek.

With the LDRT+PA framework in place we turned to de re/de se belief reports. The idea was to generate a basic relational If, i.e. a uniform de re/de se preliminary DRS where all pronouns are interpreted directly as they appear on the surface (no feature deletion). But instead of existentially quantified, the acquaintance relation is marked as underspecified: it needs to be resolved together with the sentence’s presuppositions when we combine the sentence DRS with the context. This requires second-order resolution, which we achieved by using higher-order unification. The resolution in effect outputs either a pure de se or a particular, salient de re reading, according to what fits best in the context. Under the natural assumption that equality is always the first candidate resolution, only to be overruled under severe contextual pressure, we get the observed preference for pure de se. Note how this kind of strategy is formulated very naturally in DRT+PA’s pragmatic/semantic account of resolution. The resulting system, dubbed Acquaintance Resolution, can take care of simple de re/de se reports, including Ortcutt cases and de re/pure de se ambiguities. What sets my account apart from the rival accounts of Schlenker, von Stechow and others is that I have shifted the burden of explanation onto the pragmasemantic module, keeping an entirely straightforward syntax and a unified de re/de se PrelDRS where every pronoun is interpreted according to its surface features. I believe to have shown with this thesis that the problems of belief report semantics have a more natural formulation and solution in such a semantic/pragmatic setting.

Zimmermann’s universally quantified report in a mixed scenario, and Percus and Sauerland’s only report, in 3.4.3, led us to turn the underspecified acquaintance relation into a real presupposition, thus allowing accommoda-
tion when binding fails. Complements with infinitives, logophors, and shifted indexicals formed the final obstacle (3.4.4). To overcome it, I proposed to analyze PRO, LOG and Amharic I as regular first person indexicals, i.e. as rigid presuppositions. These may be bound by the local belief center to give pure de se outputs. Technically, this move involved an extension of acquaintance to vivid concepts and the introduction of a logical principle of introspection.

4.2 Future research

Obviously, in the course of this study I have come across a number of issues that deserve further attention. In some cases the discussion was cut short because it was leading us too far off the main topic. In other cases the theory itself arguably encountered some trouble, but for reasons of space and time I could only offer a very rough sketch of an idea to improve the situation. Let us here collect some of these loose ends as pointers to further research.

As for the first chapter, one area for future research concerns the two proposals for a unified treatment of de re/de se in 1.2.3, to wit, the relational account (Kaplan 1969; Lewis 1979a; Cresswell & von Stechow 1982) and the two-dimensional belief characters of Kaplan (1989). It would be interesting to take a closer look at the exact differences between these two rather similar approaches and to see how these differences derive from the different ways they divide up the de re belief’s content into a wide (actual res) and a narrow, psychological part (cf. 1.2.4.4). Perhaps this could even shed new light on the recently observed shortcomings of Kaplan’s character theory when it comes to belief reporting (2.3.5).

In chapter 2 we’ve come across a number of issues that deserve further research. One is the analysis of abstract anaphora and ellipsis, which, as we saw in 2.4.1.4, give rise to strict/sloppy ambiguities that interact in interesting ways with the de re/de se ambiguity. The abstract anaphora/ellipsis data could then perhaps be connected more closely to the quantified reports with every and only. A related issue is my analysis of only as a quantifier, neglecting the contribution of focus and information structure. A last issue deserving of further research is the connection between quotation and attitude reports. So-called mixed quotation, discussed in 2.4.3.2, seems to fall somewhere in between indirect reporting and direct quotation, as do the Schlenkerian examples of attitude reports with shifted indexicals. It would be interesting to see if perhaps indexical shifting can be accounted for in terms of mixed quotation, or mixed quotation in terms of monstrous reporting, or that both are to be thought of as manifestations of a single underlying mechanism.
In the course of the third chapter there arose various, mostly technical, issues on which I have simply chosen an easy formalization without considering all the alternatives in detail, thus bypassing a number of technical yet interesting discussions. First, as I mentioned in 3.3, it might be possible to simplify the syntax and semantics of LDRT by getting rid of the labeling of discourse referents. The problem with this is that we would assign existentially closed interpretations to layers of LDRSs where we currently do not assign any truth-conditional interpretation, i.e. cases where an \( l \)-free occurrence of a discourse referent \( x \) (in an \( l \)-condition) has access to a universe with an \( x \), i.e. in a different layer. Furthermore, if we abandon the labeling of discourse referents, it remains to be seen if we can still talk about \( k \)- and \( fr \)-presuppositions. A second issue has to do with \( k \)-presupposition, more specifically with proper names. As of yet, I don’t know why exactly proper names seem more willing to cross over to a descriptive layer than indexicals. As I mentioned in 3.3.3.2, I follow Zimmermann’s (2004) Hypothesis L in that both are lexically specified as directly referential devices, i.e. \( k \)-presuppositions. Turning proper names into layer-unspecific anaphors is therefore not an option. As a matter of fact, this is what sets apart my account of proper names (and indexicals) from mere scopal (neo-)descriptivism as championed by Geurts (1997) and Hunter & Asher (2005). Perhaps, though, there can be a compromise along the lines of pragmatically instead of lexically assigned labels separating rigid and “at issue” content.

When we get to actual belief reporting in (L)DRT, the \textit{de dicto} mode of believing was somewhat neglected, especially the possibility of a unification of \textit{de re} and \textit{de dicto} that analyzes \textit{de dicto} belief as relational belief with ‘acquaintance’ relations to general concepts. In any case, as it stands we have brought together \textit{de re} and \textit{de se} in a single PrelDRS, but there’s quite a big gap between this and the current \textit{de dicto} PrelDRSs. The final loose end concerns my analysis of LOG and PRO on a par with Amharic and English \( I \), save for some resolution restrictions. I’ve already remarked in 3.4.4.4 that, although I think my account is unique in doing full justice to shifted \( I \) as a real semantic indexical, I find the generalization to logophors and PRO still unsatisfactory, for morphological and syntactic evidence suggest that LOG and PRO actually carry third person features. It would be worthwhile to investigate a suggestion of Schlenker’s to the effect that we might analyze those forms not as first person indexicals, but rather as the grammaticalization of the acquaintance relation of equality directly.
Glossary of names

**Essie Beard** Essie (martellax@greekchat.com) is to appear on a game show. Heavily made up and stressed out, she thinks she’s seeing her opponent while actually looking at her own mirror image. She then says: “I so hope you lose.”, 82

**Ligia Faust** Ligia Faust (pertras@mom2momlist.com) is kissed by Colby Vogel who believes her to be a philosopher, 4, 9, 34

**Avis Fish** Avis (utrtppvkzajze@home-schooled.com) is a bitchy judge on America’s Next Topmodel. She’s very confident about her own looks. One time she’s looking through a portfolio, saying “She looks awful”. Turns out, the other judges had put in some old photos of Avis herself,, 113, 114

**Ellsworth Kimmel** Ellsworth (iiicqvvi@doorworld.com) is the prettiest candidate in an unspecified contest. She also believes that the prettiest always win, though not that she herself is the prettiest. In an alternative scenario, B, she sees some ‘Arata Suggs’ on TV and becomes convinced that that so-called Arata will win. In fact, *Arata Suggs* is a nickname given to Ellsworth, and the person on TV is herself. In scene C Ellsworth is more confident and she confidently utters: “I’m gonna win this thing.”, 72, 91, 93, 119, 130, 132, 135, 219, 220, 225, 273

**Shavonne McManus** Shavonne (serafim@mickeyfan.com) is Ellsworth Kimmel’s rival. When they’re at a party together, Ellsworth tells her, “Seems Arata Suggs is going to win” (without realizing she is the one called Arata) to which Shavonne, brimming with self-confidence, replies: “Fat chance, I’m gonna win”, 132, 135, 137

**Prof. Ortcutt** Professor Ortcutt is a famous professor, senior colleague of Janell Ventress and Ligia Faust, and head of Department X. Janell admires Ortcutt a lot. With Ortcutt’s imminent retirement, however,
she hopes to succeed her as head of the department., 16, 17, 26–29, 31–34, 51, 53, 78, 86, 227, 229, 275, 285, 287, 288

**Arata Suggs** Nickname given to Ellsworth Kimmel behind her back. In scenario B, Ellsworth/Arata overhears supporters of Arata who are watching her speech on the TV, 72, 91, 93, 95, 119, 132, 219, 220, 224, 225, 275

**Janell Ventress** Janell Ventress (zorrel@religiousfollower.com) is part of a conference review committee. She believes in fair and blind reviewing. When she happens to get assigned her own submission, she doesn’t recognize it, but gives it high marks, 15–19, 24–29, 31–35, 37, 51, 52, 58, 78, 86, 227, 229, 285, 287, 288

**Colby Vogel** Colby Vogel (emaier@phil.kun.nl) kisses Ligia Faust, because he believes her to be a philosopher, 4, 9, 20
Bibliography


BIBLIOGRAPHY


Bibliography


——. 2006. Minimal pronouns: Fake indexicals as windows into the properties of bound variable pronouns. Ms. semanticsarchive.net. [3 citations on 2 pages: 141 and 164]


— —. 2004b. Person and binding (a partial survey). [draft, April 15, 2004.]. [2 citations on page 90]
Bibliography


Bibliography


Summary

John, a real Coke fan, takes the Pepsi Challenge. Before the test he is asked what his Cola of choice is. He responds as in (1a), from which the host concludes (1b).

(1)  a. John: “Coca-Cola tastes best.”
     b. John thinks that Coca-Cola tastes best

An utterance like (1b) is called a belief report. More precisely, a de dicto belief report, since it presents John as believing the proposition expressed by his own utterance in (1a). A belief report is a natural language construction meant to ascribe a belief to an individual. Though some authors fail to make a clear distinction, it clear that a theory of belief reports is something different from a theory of belief, though the first arguably presupposes the latter to some extent. So, first we have to say something about belief. Definition: John believes a proposition de dicto iff he is disposed to assent to a linguistic realization of that proposition. A de dicto belief report then is defined as a sentence of the form NP believes that S that means that the subject believes de dicto the proposition expressed by the complement, S.

John takes two sips and concludes that the drink labeled ‘M’ is the best. The host then reveals that M is actually Pepsi and declares:

(2) John thinks that Pepsi tastes best

John might disagree at first, but the facts are that he said he preferred M and that M = Pepsi, so (2) is true. This is what we call a de re belief report, since the belief it reports is not about the word or concept Pepsi, but rather about the actual object denoted by that phrase, the so-called res. What is expressed is that John thinks that this actual drink here, which happens to be known both as ‘Pepsi’ and as ‘M’, is the best. An important characteristic of de re belief is that involves a subject who is vividly acquainted with an external object, for without proper acquaintance it is impossible to have a

\footnote{http://en.wikipedia.org/wiki/Pepsi_Challenge.}
belief about an external object. This leads to the following definition of de re belief (Kaplan 1969):

\[ (3) \quad x \text{ believes de re of } y \text{ that } y \text{ is } P \text{ iff there is a relation } R \text{ with:} \]

a. \( R \) is a vivid relation of acquaintance  
b. \( R(x, y) \) in the actual world  
c. \( x \) believes the proposition that the object he is \( R \)-acquainted with is \( P \) (i.e. \( P(\tau z[R(x, z)]) \))

Applying the definition to the example, we see that John has a de re belief by taking: \( x = \text{John}; y = \text{the Pepsi in the cup labeled M}; P(x) \text{ iff } x \text{ tastes best}; \) and \( R(x, y) \text{ iff } x \text{ drinks a Cola } y \text{ from the cup marked M}.^2 \)

Besides de dicto and de re beliefs and reports, philosophers and linguists have argued for a third mode: de se (Castañeda 1966; Perry 1977; Lewis 1979a; Kaplan 1989). Consider the following real life example.\(^3\) In an experiment some male speakers of various sexual orientations were recorded after they had first reported their sexual orientation. They were then asked to rate all recorded passages, including their own, on a scale from gay to straight. Frank is gay, he knows it and deliberately tries to sound gay. When approached for the experiment he said: “I sound gay.” Frank thereby expresses a de re belief about himself, that he sounds gay. Chris, who would never describe his own voice as gay, also participates in the experiment. Now he’s listening to the recordings and when he gets to #12 he marks it as ‘very gay’, not realizing that this is his own voice. Thus, Chris also expresses a de re belief about himself, because listening carefully to recording #12 brings him into a perceptual acquaintance relation with himself. The difference between the beliefs is that Frank’s is not only de re about himself, but (pure) de se: he believes he himself sounds gay from his own first person point of view, while Chris’s belief is about himself from a third person perspective. More specifically, beliefs that the believer would express with a first person indexical (\( I \)) correspond to 1st person pure de se beliefs, or simply de se beliefs since we shall focus on the person domain exclusively in this summary.

The example brings out the importance of the indexical point of view in the analysis of belief. At a propositional level Chris’ thought “#12 sounds gay” is indistinguishable from “I sound gay”, though he might utter the

---

2Note by the way that de dicto and de re beliefs are not mutually exclusive. The de dicto belief expressed in (1b) is also de re, since we may safely assume that John has had various encounters with Coca-Cola in the (recent) past and has formed his belief on the basis of those. On the other hand (2) interpreted de dicto is false: John would never say or think “Pepsi tastes best.”

3http://itre.cis.upenn.edu/~myl/languagelog/archives/001799.html
former, but not the latter. This prompted Lewis (1979a) to change from propositions as objects of belief, to properties that the believer self-ascribes: Frank’s first person utterance means he self-ascribes the property of sounding gay, while Chris self-ascribes the distinct property of listening to recording #12, which sounds gay. We revise the last clause in the definition of de re belief accordingly:

\[ c'. \ x \text{ self-ascribes the property of being } R\text{-acquainted with a } z \text{ that is } P \ (i.e. \ x \text{ self-ascribes } \lambda z[P(\iota z[R(x,z)])]) \]

Note how de se belief comes out as a special case of de re, viz. the case where \( R \) is the relation of equality (substituting = for \( R \) above collapses the complex property to the property of being \( P \)) (Lewis 1979a:156). It is easily verified that on this definition both Frank and Chris believe de re of themselves that they sound gay, but only Frank does so under the acquaintance relation of equality, i.e. de se. Chapter 1 of this dissertation is largely an exposition and defense of the relational reduction of de se to de re belief, and of Kaplan’s (1989) comparable reduction formulated in terms of two-dimensional character semantics.

Chapter 2 is devoted to the semantics of de re and de se belief reports in natural language. We can easily make a report semantics based on the acquaintance-based, reductionist de re/de se belief semantics (Cresswell & von Stechow 1982):

\[ [\text{NP}_1 \text{ believes that } \text{NP}_2 \ \text{VP}] = 1 \text{ iff there is a relation } R \text{ with:} \]

\[ \begin{align*}
  &a. \ R \text{ is a vivid relation of acquaintance} \\
  &b. \ R([\text{NP}_1], [\text{NP}_2]) \text{ in the actual world} \\
  &c. \ [\text{NP}_1] \text{ self-ascribes the property of being } R\text{-acquainted with a } z \text{ that is } [\text{VP}] 
\end{align*} \]

In cases where \( \text{NP}_1 \) (the subject) and \( \text{NP}_2 \) (the res) are co-referential, we predict that the reported belief could be de se or (merely) de re. In other words, since a report like (5) is linguistically underspecified for de re or de se:

\[ \text{(5)} \quad \text{Frank thinks he sounds gay} \]

Applying (4) gives that (5) is true iff there is a vivid relation of acquaintance between Frank and himself such that Frank self-ascribes being so acquainted with a unique person that is gay. And this is true because Frank believes himself to be gay, so we can take the relation of equality. But we would also predict that (6) is true:
(6) Chris thinks he sounds gay

Since Chris doesn’t even recognize his own voice, this prediction may seem problematic at first, but I follow the (philosophers’) consensus that it is nonetheless correct. Admittedly, it would be considered misleading to word Chris’ predicament in this way if the circumstances of his mistaken identity weren’t already mentioned or otherwise firmly established in the context (e.g. if we qualify the report by adding “…, but he doesn’t realize it”, or visually, as when the reporter and her audience are watching the whole mistaken identity scene unfold).

One of the main selling points of this relational semantics is that for sentences like (6) it can do without postulating any kind of ambiguity as a linguistic correlate to the de re/de se distinction. The flip side is that such unified theories cannot explain the obvious difference in acceptability between (5) and (6) in the given context, even if they are both strictly speaking true. A more serious drawback of unification is that we’re at a loss to deal with reports that can only be read de se, as, for example, the English infinitival complement construction:

(7) a. Frank believes to have a gay voice

b. #Chris believes to have a gay voice [*though he doesn’t realize it]

Chierchia (1989) uses this contrast to argue for a dedicated de se LF, which means that the construction in (6) is ambiguous between a de re and de se LF while (7) isn’t. This ‘ambiguity thesis’ is compatible with syntactic theory telling us that the reports in (6) and (7) share the same form, ‘NP believes [NP VP]’, with the second NP slot filled by third person pronoun in (6), and by an unpronounced PRO in (7). Chierchia analyzes this PRO as a de se pronoun, i.e. an element that has to be bound by the belief center, forcing the report into a de se LF, while an overt pronoun is only optionally bound de se. Schlenker (2003) later argues that the so-called logophoric pronouns found in many African languages can also be seen as de se pronouns.

---

4The same is true for Kaplan’s quantified character analysis, but von Stechow & Zimmermann (2004) lay bare a fatal flaw in that account, as I show in my thesis, so I’ll disregard it here.

5Terminology: ‘#’ denotes the semantic judgment that the sentence uttered in a given context is false; ‘*’ denotes the syntactic judgment that a sentence (or continuation) is ungrammatical.

6The so-called hybrid variant lets (5) and (6) keep their unified relational LF while the reports in (7) receive a totally different syntactic and semantic analysis. The main argument against this is that it does not extend to the arguably related phenomena of logophoricity and shifted indexicality (see below).
Another argument for the ambiguity thesis (against unification) is due to Percus & Sauerland (2003a) and involves the following quantified report:

(8) Only Frank thinks he sounds gay

This report is commonly judged true in our mixed de re and de se scenario. On a relational account it would come out false, because Frank is not the only one who believes of himself, under some acquaintance relation, that he sounds gay; he’s just the only one who has such a belief under the acquaintance relation of equality. Note that the simple fix of giving the quantified acquaintance relation wide scope over the overt quantifier only will wrongly predict that (9a) is true and (9b) is false:

(9) a. #Only Chris thinks he sounds gay
    b. Both guys think they sound gay

A final argument against the unified relational analysis originates in Schlenker’s (2003) work on shifted indexicals. In many languages (including Amharic, Ancient Greek and Zazaki), an unmistakably indexical first person form embedded in a belief or speech report may refer not to the actual speaker of the report, but to the subject of the attitude:

(10) jon ṭən ga-na na -ūn yil -all  
     john hero be -1.sg say.3.sg -aux.3.sg  
     ‘John says that he is a hero’

Schlenker builds a framework in which this kind of construction is semantically speaking the default way to report a de se attitude, but on the relational account such a shifted first person is highly problematic. This is because de se is then a relational, de re attitude. Applying (4) to (10) with NP$_1$=John and NP$_2$=1.sg, it is obvious that the acquaintance relation can never be identity, because John is not the actual speaker, so we incorrectly predict that (10) means the same as its literal English counterpart: John says that I am a hero.

Despite all these apparent arguments against unification, I propose a new unified semantics of de re/de se reports, built on the relational attitudes framework, but incorporating the context-dependence of acquaintance. The resulting theory, Acquaintance Resolution, handles a wide variety of data including the Pepsi Challenge, de re/de se ambiguities, infinitival complements, logophors, quantified reports and shifted indexicals. Thus, the problems raised above are all solved, and we achieve a unification of de re and de se attitude reports.
The proposed account is based on the idea of acquaintance relations being provided by the context rather than existentially quantified over. Taking this idea seriously, I decided to develop my theory in a dynamic, representational semantic framework: Discourse Representation Theory (DRT), a framework designed to accommodate context-dependence. To account for the kind of phenomena we’re interested in here, I propose a few extensions to basic DRT. The first is standard, I incorporate van der Sandt’s (1992) theory of Presupposition as Anaphora to analyze pronouns and other definites uniformly. With that comes a two-stage interpretation architecture: from the syntactic analysis of a sentence a preliminary DRS is constructed fully compositionally, and then a pragmatic/semantic resolution mechanism connects that preliminary structure with the context DRS by resolving its presuppositions. Acquaintance Resolution will be compositional and unified in the sense that the preliminary DRS is generated compositionally, and underspecified for \textit{de re}/\textit{de se}. The final output DRS is of course determined in part pragmatically, and in interaction with the context: it is neither fully compositional (in the strong, Montagovian sense), nor unified, since a typical \textit{de re} acquaintance and a \textit{de se} one give distinct truth-conditions.

Second, I apply a new version of Layered DRT (Geurts & Maier 2003) to divide DRSs into two separate but interconnected layers. LDRT was originally designed as a general framework for representing different types of information within a single DRS. All conditions reside in a specific layer, but these layers can all use the same discourse referents. I use it here to separate descriptive, at issue content, from rigid, directly referential content. Combined with the analysis of definites as presuppositions this leads to a powerful new semantics of proper names and indexicals. The relevance to the current undertaking lies primarily with the indexicals. In particular, LDRT gives, to my knowledge, the first principled semantics of the first person pronoun in DRT. A final addition to the standard DRT framework is the switch from believed propositions to self-ascribed properties, which is analogous to that in the classical framework.

With these preliminaries in place we turn to Acquaintance Resolution proper. The idea of a context-dependent acquaintance relation can be formalized as a kind of presupposition. The preliminary DRS contains a free relation variable $R$ that stands for the acquaintance relation and thus for the description under which the belief is held. In the resolution stage, this $R$ is resolved by binding it to a suitable relation already present in the context DRS. This binding is subject to the usual pragmatic and semantic presupposition binding constraints (salience, accessibility, consistency etc.) and some new ones: (i) the antecedent relation should hold between the report’s believer and $res$, (ii) it should be a proper relation of acquaintance (e.g. a perceptual
relation), and (iii) resolving to equality is the default. The last clause ensures a preference for de se readings in co-referential reports, thus answering the first objection against the relational framework mentioned above. Moreover, it explains why people hesitate to accept (6) as a true report of our mistaken identity scenario, because we first resolve to equality, which leads to an incoherent output DRS, after which we have to recompute and search the context for a suitable acquaintance relation. The first additional constraint, (i), will be built into the second-order binding mechanism, which I formalize in terms of higher-order unification (Huet 1975; Dalrymple et al. 1991). This means that we equate the DRS condition that \( R \) holds between believer and res with a ‘suitably parallel’ condition from the context, and then find a unifying substitution for this (higher-order) equation. The resolution consists in applying this substitution (for \( R \)) to the whole DRS.

Let’s illustrate all this with an example resolution of (6). First, we construct the preliminary DRS for the ambiguous (6), repeated below. Note that the bel operator denotes the operation of property self-ascription, so the embedded belief DRS denotes a set of contexts, of which center picks out the agent/speaker parameter:

(11) Chris thinks he sounds gay

\[
\begin{array}{c}
\begin{array}{c}
R(z,w) \\
\hline
z \quad w \\
\hline
\text{chris(z) \ 3.sg.m(w)} \\
\hline
\end{array} \\
\hline
\end{array}
\]

\[
\begin{array}{c}
\begin{array}{c}
\text{BEL}_z \\
\hline
\text{center(u)} \\
\text{R(u,v)} \\
\text{sounds_gay(v)} \\
\end{array} \\
\hline
\end{array}
\]

The free R represents the underspecified acquaintance relation, the dashed boxes represent the presuppositions triggered by the subject (Chris) en res-term (he) of the report. Next, we add this preliminary DRS to a context, in this case a DRS where it is salient (to the reporter and her audience) that Chris is hearing his own voice. After that we bind the two presuppositions:
The last step is binding $R$ by determining a substitution that unifies $R(x, x)$ with the contextually salient $\text{hear}(x, x)$:

$$
\text{(13)} \quad \leadsto R(x, x) \doteq \text{hear}(x, x) \leadsto [R \leadsto \lambda s \lambda t [\text{hear}(s, t)]]
$$

In the resulting output Chris ($x$) thinks he ($u$) is hearing someone ($v$) who sounds gay. Note that we only chose this resolution path because choosing equality would have led to an output with $u = v$, a de se report requiring that Chris thought “I sound gay”, which is inconsistent with the story. If it had been Frank instead of Chris, that reading would have been possible and thus preferred.

This basic framework is extended in two steps to account for all problematic report data discussed above. The first extension allows us to handle the quantified data. As it stands, embedding a pronoun report like (5) under a quantifier will only generate quantified de se readings, i.e. Both guys think they sound gay is only true if both Chris and Frank have the right de se attitude. This is because the subject parameter in the acquaintance relation is a quantified variable, which does not figure in any contextual condition, so we can only take the default de se resolution. As (9b) shows, this prediction is not borne out. In order to solve this problem, we make the resolution of $R$ more flexible by allowing projection and accommodation, in other words: make it a real second-order presupposition. Zimmermann’s
judgment comes out as a case of local accommodation of $R$. Again, this option is only available because the default $de se$ resolution is inconsistent with the mixed context. Percus & Sauerland’s (2003a) judgment about (8) corresponds rather straightforwardly to the default resolution (given that $only$ is still treated as a quantifier). With Zeevat’s (9a) indeed all possible resolution options (binding and accommodation, global and non-global) fail to produce a true output, mainly because the $R$-presupposition is trapped inside the duplex condition representing the quantifier. So this problem is solved as well.

The second and last extension is somewhat more involved. It concerns the unambiguous $de se$ reports with infinitives and logophors, and the shifted first person indexicals. My solution starts with the latter and applies that to the former. Basically, the reason Acquaintance Resolution as it stands has trouble with shifted $I$ is precisely because it treats $de se$ as a subspecies of $de re$. The $res$ is logically represented outside the belief context, so a first person pronoun syntactically embedded as $res$ will always refer to the actual speaker. This corresponds to Kaplan’s (1989) well-known ‘prohibition of monsters’, which works fine for English $he$ and $I$, but fails for Amharic and other languages that allow shifted indexicals. The first step towards a solution is to generate the embedded subject presupposition in situ, i.e. within the belief box of the preliminary DRS, and rely on the general preference for wide scope resolution, aided by the $k$ layering, to ensure that referential terms and other definites end up in the main context. This leaves open the possibility of a local binding, and that is exactly what happens with the shifted reading of an Amharic belief embedded $I$: the belief center, sharing the first person content and the $k$ layer of indexicality, binds the $res$ presupposition:

(14) John says that I am a hero

\[
\begin{array}{c|c}
\hline
x & \hline
john(x) & \hline
R(x, w) & \hline
\end{array}
\]

\[
\begin{array}{c|c}
\hline
x & \hline
\sim & \hline
\end{array}
\]

\[
\begin{array}{c|c}
\hline
u & v \hline
\end{array}
\]

\[
\begin{array}{c|c}
\hline
BEL_x & \hline
center(u) & \hline
R(u, v) & \hline
hero(v) & \hline
\end{array}
\]

\[[w \mapsto u]\]

\[
\begin{array}{c|c}
\hline
x & \hline
john(x) & \hline
R(x, u) & \hline
\end{array}
\]

\[
\begin{array}{c|c}
\hline
u & v \hline
\end{array}
\]

\[
\begin{array}{c|c}
\hline
BEL_x & \hline
center(u) & \hline
R(u, v) & \hline
hero(v) & \hline
\end{array}
\]

\[[1,s g(w)]\]
From here we need to eventually arrive at a *de se* output. But, since $x$ is not equal to $u$ (not to mention the fact that this $u$ occurs here as a free variable), how can we bind $R$ to equality? A closer look reveals that $x$ may not be equal to $u$, but they are rather intimately connected: $u$ is the center of $x$’s belief. The next step is to generalize the notion of acquaintance to allow for the possibility of acquaintance to concepts (intensional discourse referents). If we do that, we can safely close off the free (now intensional) variable in the main context when $x$ is acquainted with his self-concept, the function mapping each of his belief alternatives to its center, in other words: the person he believes to be. In order to uphold a *de dicto/de re* distinction, we must be cautious in admitting *de re* acquaintance with a mere concept, but it seems reasonable to say that the self-concept at least is a viable candidate *res*. Binding to his acquaintance with the self-concept gives a DRS in which John’s belief center’s belief center is believed to be a hero. I invoke a rather uncontroversial introspection principle from doxastic logic to reduce the double belief embedding, by which we derive the proper *de se* reading. It’s now also obvious what to do with PRO and logophors: represent them semantically as first person pronouns.

The end result is still a truly unified analysis of *de re* and *de se* reports, in the sense that we generate fully compositional but highly underspecified preliminary structures. I’ve collected the evidence against *de re/de se* unification, and showed how my account can deal with all the alleged counterexamples and more. The account uses a couple of independently motivated semantic and pragmatic tools (e.g. higher-order unification, presupposition resolution, two-dimensional semantics, and introspection) while keeping the syntax entirely straightforward (e.g. no movements, (almost) no feature deletion under agreement, and (almost) no invisible items). One of the extended aims of this work was to interpret all pronouns and other lexical items in a report as they appeared on the surface. I’ll let the reader decide if I have succeeded in that and thus made a convincing case for the unification of *de re* and *de se* through Acquaintance Resolution.
Geloof in Context

Samenvatting

Dit proefschrift gaat over attitude toeschrijvingen. Meer in het bijzonder zal ik een uniforme semantiek van de re en de se geloofstoeschrijvingen geven. Een de re geloof is een geloof dat een individu heeft over een bepaald extern object in haar onmiddellijke omgeving, bijvoorbeeld mijn geloof dat deze trein waar ik nu in zit naar Amsterdam gaat. Ik denk namelijk dat ik in de trein van 10:50 van Nijmegen naar Amsterdam en verder zit. Maar stel nu eens dat ik per ongeluk in the 10:56 naar Zwolle ben gestapt. Als ik nu tegen een onzekere medereiziger zou zeggen: “Deze trein gaat naar Amsterdam”, dan zou ik daarmee dus eigenlijk een de re geloof uitdrukken over de 10:56 trein, nl. dat hij naar Amsterdam gaat. En dit ondanks het feit dat ik zelf nooit zou zeggen: “De trein van 10:56 gaat naar Amsterdam”. Dit is een van de fundamentele puzzels met betrekking tot de re geloof, en een die in dit proefschrift ook een grote rol speelt.

De se geloof is gedefinieerd als iets wat de gelover zelf met context afhankelijke uitdrukkingen zoals ik, jij, hier onder woorden kan brengen. Bijvoorbeeld Jan’s geloof dat hij er goed uit ziet, door hemzelf uitgedrukt als “ik zie er nog goed uit”. De uitdrukking ik is essentieel voor dit geloof, het is immers niet hetzelfde als het geloof dat Jan uitdrukt met “Jan ziet er nog goed uit”, aangezien Jan best zijn naam kan zijn vergeten, bijvoorbeeld door geheugenverlies, en dan met beide zinnen duidelijk iets anders bedoelt. En idem als hij zou zeggen “die ziet er nog goed uit” wijzend naar een live video opname van zichzelf. Hij drukt daarmee weliswaar dezelfde propositie uit, dat Jan er nog goed uit ziet, maar het onderliggende geloof is anders, want hij zou het laatste bijvoorbeeld kunnen denken zonder zichzelf op het scherm te herkennen en dus zonder meteen te denken “ik zie er nog goed uit”. In alle drie de gevallen drukt Jan een de re geloof uit over zichzelf, maar alleen die met ik noemen we (puur) de se.

In hoofdstuk 1 laat ik zien dat de se inderdaad altijd gezien kan worden als een speciaal geval van de re. Vervolgens richten we ons in hoofdstuk
2 en 3 op geloofstoeschrijvingen. Dat zijn zinnen als “Emar denkt dat deze trein naar Amsterdam gaat”, die zeggen dat iemand een bepaald geloof heeft. Een *de re* of *de se* toeschrijving is dan een zin die zegt dat iemand een *de re* respectievelijk *de se* geloof heeft. Mijn centrale stelling is dat de reductie van *de se* als een speciaal geval van *de re* ook voor toeschrijvingen geldt. Er is echter zeker geen ‘een op ‘e’en correspondentie tussen *de re* en *de se* geloof en geloofstoeschrijvingen in natuurlijke taal. Veel zinnen zijn bijvoorbeeld ambigu of onderspecificeerd wat betreft het geloofstype:

(1) Jan denkt dat hij er nog goed uitziet

Deze zin is natuurlijk waar als Jan denkt “ik zie er nog goed uit” (*de se*), maar, met enige moeite willen we ook nog wel toegeven dat hij waar is in een situatie waarin we zien dat Jan denkt “hij daar op die video ziet er goed uit” zonder het door te hebben kijkend naar zichzelf (*de re*).

Dit volgt mooi uit de analyse van *de se* als *de re* maar er zijn ook toeschrijvingen die alleen *de se* te lezen zijn:

(2) Jan denkt er nog goed uit te zien

Deze zin is duidelijk onwaar als Jan denkt dat hij het over iemand anders heeft. En er zijn nog meer constructies in natuurlijke talen die tegen de uniforme behandeling van *de re* en *de se* spreken. Zin (3) bijvoorbeeld lijkt zoiets te betekenen als: Jan is de enige die denkt, “ik zie er nog goed uit”:

(3) Alleen Jan denkt dat hij er nog goed uit ziet

In andere talen komen we nog meer toeschrijvingen tegen die uitsluitend *de se* te lezen zijn, bijvoorbeeld de zogenaamde logophoren in veel Afrikaanse talen. In weer andere talen vinden we toeschrijvingen met een ingebedde *ik* die niet naar de feitelijke spreker verwijst (‘monstrously shifted indexicals’ in o.a. het Amhaars, Zazaki en Oud-Grieks), hetgeen eveneens in strijd is met de uniforme *de re/de se* analyses. In hoofdstuk 2 bespreek ik al deze data en andere argumenten tegen de uniforme benadering van *de re* en *de se* toeschrijvingen.

Tenslotte laat ik in hoofdstuk 3 zien hoe we de uniforme *de re/de se* analyse kunnen redden: door over te stappen op een dynamische semantiek. Ik ontwikkel een systeem waarin uit een toeschrijving onderspecificeerde logische vormen worden gegenereerd, waarvan de context bepaalt wat voor *de re* geloof het precies wordt, met *de se* als default. Dat basis systeem maakt gebruik van enkele onafhankelijk gemotiveerde semantisch-pragmatische tools.

---

1Voeg bijvoorbeeld “…, maar hij heeft het zelf niet door!” toe aan (1).
zoals presuppositions, een uitbreiding van DRT (‘Layered DRT’) met tweedimensionale semantiek, en tweede orde unificatie. Dan volgen nog twee technische uitbreidingen om echt alle eerder besproken problemen tegen de unificatie te weerleggen. Het eind resultaat is een uniforme, dynamische analyse die recht doet aan een zeer breed scala aan *de re* en *de se* puzzels en probleemgevallen voor eerdere analyses.
Curriculum Vitæ


From 2001 to early 2006 he worked as PhD student at the Philosophy Department, partly in the NWO project *Information Integration in Discourse*. Supervised by prof. dr. Rob van der Sandt and dr. Bart Geurts, he has published and presented research on various subjects in the semantics/pragmatics interface, including attitude reports, direct reference, indexicality, DRT, presupposition, quotation, contrast, denial, and multi-dimensional semantics. He is currently teaching a logic course at the University of Amsterdam.